Voice Outcomes after Total Thyroidectomy, Partial Thyroidectomy, or Non-Neck Surgery Using a Prospective Multifactorial Assessment

Diego A Vicente, MD, Nancy P Solomon, PhD, CCC-SLP, Itzhak Avital, MD, FACS, Leonard R Henry, MD, FACS, Robin S Howard, MA, Leah B Helou, MA, CCC-SLP, George L Coppit, MD, Craig D Shriver, MD, FACS, Chester C Buckenmaier, MD, Steven K Libutti, MD, FACS, Ashok R Shaha, MD, FACS, Alexander Stojadinovic, MD, FACS

BACKGROUND: Voice alteration remains a significant complication of thyroid surgery. We present a comparison of voice outcomes between total thyroidectomy (TT), partial thyroidectomy (PT), and non-neck (NN) surgery using a multifactorial voice-outcomes classification tool.

STUDY DESIGN: Patients with normal voice (n = 112) were enrolled between July 2004 and March 2009. The patients underwent TT (n = 54), PT (n = 35), or NN (n = 23) surgery under general endotracheal anesthesia as part of a prospective observational study involving serial multimodality voice evaluation preoperatively, and at 2 weeks, 3 months, and 6 months postoperatively. Patients with adverse voice outcomes were grouped into the negative voice outcomes (NegVO) category, including patients with objective (abnormality on videolaryngostroboscopy and substantial voice dysfunction) and subjective (normal videolaryngostroboscopy but with notable voice impairment) NegVO. Voice outcomes were compared among study groups.

RESULTS: Negative voice outcomes occurred in 46% (95% CI, 34–59%) and 14% (95% CI, 6–30%) of TT and PT groups, respectively. No NegVOs were observed after NN surgery. Early NegVOs were more common in the TT group than in the NN or PT groups (p < 0.001). Most voice disturbances resolved by 6 months (TT 84%; PT 92%) with no difference in NegVO among all groups (p = 0.23). Black race and significant changes in certain voice outcomes measures at the 2-week follow-up visit were identified as predictors of late (3 to 6 months) NegVO.

CONCLUSIONS: This comprehensive voice outcomes study revealed that the extent of thyroidectomy impacts voice outcomes in the early postoperative period, and identified risk factors for late NegVO in post-thyroidectomy patients who should be considered for early voice rehabilitation referral. (J Am Coll Surg 2014;219:152–163. © 2014 by the American College of Surgeons)
An estimated 56,460 new cases of thyroid cancer were projected to be diagnosed in 2012.1 In addition, millions of patients in the United States are diagnosed each year with endocrine or inflammatory disorders of the thyroid.2 Of these patients, >80,000 will undergo thyroid surgery,3 which is reportedly associated with a 25% to 84% risk of postoperative voice alterations.4-10

Traditionally, these voice changes were attributed to the complication of iatrogenic recurrent laryngeal nerve (RLN) injury. However, most thyroid surgeons routinely either encounter or dissect the RLN during operation to preserve the nerve’s delicate structure and function.11 Adherence to meticulous surgical technique translates into low overall incidence of permanent RLN injury (0.4% to 2.5%) in experienced centers.12-14

In cases where RLN dysfunction does occur, up to 50% of these patients can have a subjectively normal voice despite abnormal objective findings on videolaryngostroboscopy (VLS).15 The presence or absence of RLN dysfunction does not solely predict functional voice outcomes after thyroidectomy.16,17 Importantly, other mechanisms can affect vocal function, including injury to the external branch of the superior laryngeal nerve,18-25 postoperative inflammation, vascular congestion and laryngeal edema, surgical trauma to the cricothyroid muscle or cricoarytenoid joint, endotracheal intubation-related trauma, and laryngotracheal fixation by postoperative fibrosis about the operative thyroid bed.6,8,18-29 Vocal manifestations from these various pathologies can range from a seemingly normal voice, to transient voice fatigue, to profound and permanent dysphonia with a substantially adverse impact on the patient’s quality of life, particularly for voice professionals.6,10,17,27

Although most post-thyroidectomy voice changes resolve spontaneously within 3 to 6 months of thyroid surgery,28,29 patients can develop maladaptive compensatory mechanisms during post-operative recovery. Such vocal behaviors can persist after resolution of the underlying vocal pathology, and are most appropriately evaluated and treated by experienced speech-language pathologists. Timely recognition of voice dysfunction and referral to a speech-language pathologist for vocal rehabilitation could be beneficial to these patients so that vocal function can be optimized.30 Therefore, post-thyroidectomy voice assessment should be initiated early and include comprehensive voice-specific functional and physical evaluations.6,24 To this end, our group has developed a longitudinal dimensional voice outcomes classification tool that incorporates patient interview, questionnaire, auditory and visual evaluation, acoustic measures, and laryngeal imaging.31-33

The literature describes various methods for evaluating postoperative dysphonia,34 and reports mixed results about how the extent of thyroidectomy affects postoperative dysphonia. In this study, we aimed to evaluate the effect of the extent of thyroidectomy on voice outcomes using our multidimensional voice outcomes classification tool. Voice outcomes were compared according to nature and extent of surgery (ie, total thyroidectomy [TT], partial thyroidectomy [PT], and non-neck surgery [NN]) performed under general endotracheal anesthesia as a basis for comparison among all groups. In addition, we evaluated thyroidectomy patient characteristics and early postoperative (2 weeks) voice outcomes measures to identify predictors of late postoperative (3 to 6 months) adverse voice outcomes, and created a practical clinical decision algorithm for the management of postoperative dysphonia readily available to thyroid surgeons.

**METHODS**

**Participants**

From July 2004 to March 2009 consenting adults undergoing thyroidectomy, parathyroidectomy, or NN at the Walter Reed Army Medical Center were enrolled in the Institutional Review Board—approved prospective longitudinal voice outcomes study (referred to as the Voice Outcomes Appraisal, Longitudinal [VOCAL] Trial). Patients were excluded from enrollment if they had a history of voice disorders before operation; earlier neck surgery; preoperative voice dysfunction demonstrated on VLS or by speech-language pathologist evaluation. For this subgroup analysis comparing voice outcomes with respect to extent of thyroidectomy in patients undergoing general endotracheal anesthesia, additional exclusion criteria included patients undergoing non-thyroid neck surgery (ie, parathyroidectomy) or surgery under regional anesthesia without general endotracheal anesthesia. The remaining patients were then divided into 2

<table>
<thead>
<tr>
<th>Abbreviations and Acronyms</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAPE-V</td>
</tr>
<tr>
<td>DSI</td>
</tr>
<tr>
<td>NegVO</td>
</tr>
<tr>
<td>NegVOobj</td>
</tr>
<tr>
<td>NegVOsubj</td>
</tr>
<tr>
<td>NN</td>
</tr>
<tr>
<td>PT</td>
</tr>
<tr>
<td>RLN</td>
</tr>
<tr>
<td>ROC</td>
</tr>
<tr>
<td>TT</td>
</tr>
<tr>
<td>VHI</td>
</tr>
<tr>
<td>VLS</td>
</tr>
<tr>
<td>VOCAL</td>
</tr>
</tbody>
</table>

---

**Abbreviations and Acronyms**

- **CAPE-V**: Consensus Auditory-Perceptual Evaluation of Voice
- **DSI**: Dysphonia Severity Index
- **NegVO**: negative voice outcomes
- **NegVOobj**: objective negative voice outcomes
- **NegVOsubj**: subjective negative voice outcomes
- **NN**: non-neck surgery
- **PT**: partial thyroidectomy
- **RLN**: recurrent laryngeal nerve
- **ROC**: receiver operating characteristic
- **TT**: total thyroidectomy
- **VHI**: Voice Handicap Index
- **VLS**: Videolaryngostroboscopy
- **VOCAL**: Voice Outcomes Appraisal, Longitudinal
thyroidectomy groups for analysis—TT and PT—and a third group of patients undergoing NN under general endotracheal anesthesia, served as a control group.

**Surgery**

All operations were performed at Walter Reed Army Medical Center, a university-affiliated tertiary referral center for military health care beneficiaries. Operations included total thyroidectomies, thyroid lobectomy and isthmusectomy (partial thyroidectomies), and NNs, all under general endotracheal anesthesia. Given the volume to outcomes relationship associated with thyroidectomies, all neck operations were performed by experienced endocrine surgeons (n = 10). Non-neck operations, performed by board-certified attending general surgeons, included laparoscopic cholecystectomy, bariatric procedures, or hernia repair.

**Participant characteristics**

Demographic information and pathology results were recorded on case report forms. Demographic characteristics that have been shown to impact postoperative voice measures include age, sex, race, BMI, smoking status, and active voice professional (e.g., singer, teacher, coach). Pathology results were noted as well, given an increased risk of postoperative RLN dysfunction after operations for inflammatory and malignant thyroid diseases.

The use of laryngeal neuromonitoring, division of the sternothyroid muscle, presence of multinodular goiter, and postoperative thyroid-stimulating hormone levels have previously been demonstrated not to be significant determinants in our voice outcomes classification tool, and were not considered in this analysis.

**Voice outcomes measures**

The voice outcomes measures used in this analysis of VOCAL trial patients have been described in detail previously and are described in Table 1. In brief, patients underwent a comprehensive multidimensional voice assessment preoperatively and postoperatively at the 2 week, 3-month, and 6-month follow-up visits. At each visit, patients completed voice case history and Voice Handicap Index (VHI) forms, recorded standard vocal tasks for expert perceptual ratings (Consensus Auditory-Perceptual Evaluation of Voice [CAPE-V]) and for calculation of Dysphonia Severity Index (DSI) and underwent VLS examination.

Patients reported voice symptoms on the voice case history. Responses that postoperative voice was “atypical” compared with preoperative voice were considered to reflect notable voice alteration. The VHI, a 30-item questionnaire with a maximum (severe) score of 120, indicates the impact of voice alterations on quality of life. For the purposes of this study, an increase in VHI score of ≥20 was established as a measure of clinically relevant voice dysfunction.

Voice samples were recorded in a double-walled sound-attenuating booth with the Computerized Speech Lab (CSL 4500; KayPENTAX) and high-fidelity microphone (AKG C420; AKG Acoustics). Tasks included standard CAPE-V vowels and sentences, highest and lowest pitch for sustained vowel sounds at various intensity levels, and maximum phonation time. Three certified speech-language pathologists independently blind-rated the CAPE-V samples for overall severity of dysphonia; an increase in the median score of ≥11% from baseline was considered clinically significant. The DSI was calculated from vocal jitter, maximum phonation time, greatest vocal fundamental frequency, and least intense phonation. A decrease in DSI of ≥2.49 from baseline was considered clinically meaningful.

Videolaryngostroboscopy was performed during each study visit with a flexible fiberoptic endoscopic nasolaryngoscope (Pentax FNL-10RP3) to evaluate the structure and function of the larynx using a variety of phonatory and nonphonatory tasks. Diagnoses of phonatory abnormalities were confirmed through consensus determination by a team of speech-language pathologists and thyroid surgeons.

Additional VOCAL trial aerodynamic and acoustic assessments were not included in this investigation, given that aerodynamic evaluation was not sensitive to post-thyroidectomy voice outcomes and cepstral/spectral evaluation, although effective in predicting voice outcomes, is currently not readily available for practical widespread clinical use.

**Voice outcomes**

The voice outcomes classification tool used in this study is illustrated in Figure 1. It was developed by a team of speech-language pathologists, surgical oncologists, and otorhinolaryngologists, and uses certain independently verified measures, and has been implemented to evaluate the impact of neuromonitoring, race, and acoustic analysis on postoperative voice outcomes.

As seen in Figure 1, patients with significant functional vocal alteration in at least 2 postoperative outcomes measures were classified as having negative voice outcomes (NegVO). These patients were then characterized as objective negative voice outcomes (NegVOobj) if they had a VLS abnormality and at least one other significant voice outcomes alteration. Patients with no visible anatomic or physiologic abnormality identified on VLS,
and 2 or more voice outcomes alterations were categorized with subjective negative voice outcomes (NegVOsubj). The remaining patients were classified as normal voice outcomes.

**STATISTICAL ANALYSIS**

Continuous data that met assumptions for normality by the Shapiro-Wilk test are presented as mean ± SD, and were compared among the TT, PT, and NN groups using 1-way analysis of variance (or the 2-sample t-test for TT vs PT). Data that were not normally distributed are presented as median with interquartile range and groups were compared using Kruskal-Wallis analysis of variance or the Wilcoxon rank-sum test. Categorical data are presented as counts with proportions (with 95% CIs for proportions computed using the modified Wald method) and compared among groups using Fisher’s exact test. Changes in voice outcomes measures and voice outcomes over time were examined by logistic regression using generalized estimating equations.

Postoperative voice outcomes were classified as early (2 weeks) and late (3 and 6 month) NegVO to differentiate patients with transient voice alterations and those at risk for persistent dysphonia and development of maladaptive compensatory mechanisms. For thyroidectomy patients (TT and PT), the association of demographic and clinical characteristics with transient and with late outcomes was examined using Fisher’s exact test. Odds ratios (ORs) and

<table>
<thead>
<tr>
<th>Voice outcomes measure</th>
<th>Description of voice outcomes measure</th>
<th>Significant change in voice outcomes measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voice case history</td>
<td>Perception of voice alteration</td>
<td>Atypical voice</td>
</tr>
<tr>
<td>Voice Handicap Index</td>
<td>Perceived impact of voice alteration on quality of life on a 120-point scale</td>
<td>Increase ≥ 20</td>
</tr>
<tr>
<td>Consensus Auditory-Perceptual Evaluation of Voice</td>
<td>Speech language pathologist evaluation of dysphonia on a 100-point scale</td>
<td>Increase ≥ 11</td>
</tr>
<tr>
<td>Dysphonia Severity Index</td>
<td>Calculation of change between pre- and post-operative acoustic measures</td>
<td>Decrease ≥ 2.49</td>
</tr>
<tr>
<td>Videolaryngostroboscopy</td>
<td>Fiberoptic evaluation of larynx to evaluate for physiologic or anatomic abnormalities</td>
<td></td>
</tr>
</tbody>
</table>

**Table 1. Voice Outcomes Measures Used in the Voice Outcomes Appraisal, Longitudinal Trial Voice Outcomes Classification Tool**

**Figure 1.** Voice Outcomes Appraisal, Longitudinal trial voice outcomes classification tool per significant voice outcomes measures. *CAPE-V, Consensus Auditory-Perceptual Evaluation of Voice; DSI, Dysphonia Severity Index; NegVOobj, objective negative voice outcomes; NegVOsubj, subjective negative voice outcomes; VLS, videolaryngostroboscopy; VHI, Voice Handicap Index.
negative and positive predictive value with 95% CI are presented.

To develop a clinical algorithm to predict late NegVO, the association between the 2-week postoperative voice outcomes measures and late NegVO was examined. Receiver operating characteristic (ROC) curves describe the diagnostic use of 2-week voice outcomes measures in predicting late NegVO. The 2-week VHI score was included in this ROC analysis because it would be a practical and easily obtainable value to quantify voice alteration in a patient with suspected post-thyroidectomy dysphonia who did not undergo preoperative voice evaluation. Data were analyzed using SPSS for Windows (version 21, SPSS/IBM).

RESULTS

Participant flow

Figure 2 depicts a flow diagram of patient participation in the VOCAL Trial from enrollment through the 6-month postoperative visit for this analysis of patients undergoing TT, PT, or NN under general endotracheal anesthesia. Of the 132 patients initially enrolled into the VOCAL Trial, 112 met all inclusion criteria for this analysis and underwent voice outcomes evaluation.

Patient characteristics

Baseline characteristics, listed in Table 2, demonstrate that the NN group had higher BMI and more white patients compared with other groups, but otherwise the 3 groups’ demographic characteristics were statistically similar. The TT and PT groups were similar in all baseline characteristics. Compared with the PT group, the TT group expectedly had a larger volume of resected thyroid gland (142 cm³ vs 32 cm³; p < 0.001) and more malignancies diagnosed (TT 54% vs PT 20%; p = 0.001). Incidence of inflammatory disease was not statistically different between thyroidectomy groups. One patient in the TT group had a documented RLN transection during mobilization of a large posterior multinodular goiter from the superior mediastinum. Preoperative voice assessment was similar across groups.

Voice outcomes measures and voice outcomes according to extent of thyroidectomy

Analysis of changes in voice outcomes measures during the postoperative period within each study group, as seen in Table 3, revealed significant changes from baseline (indicated by delta [Δ]) for all voice parameters during the 6-month follow-up period in the TT group (voice case history, p = 0.008; ΔVHI, p < 0.001; ΔCAPE-V, p = 0.014; ΔDSI, p = 0.004; VLS, p = 0.037). The number of patients classified as NegVO in the PT or NN group did not change significantly over time. Similarly, as seen in Table 4, the TT group had significant improvement in voice outcomes over time, and the PT group had better voice outcomes immediately and during the course of the study period. There were no negative voice outcomes in the NN group.
Negative voice outcomes were common in the early postoperative period after thyroidectomy, irrespective of the extent of thyroid resection (TT 43%; 95% CI, 34–59% and PT 11%; 95% CI, 6–30%). As seen in Figure 3, the TT group had significantly more NegVO \( (p = 0.002) \), NegVOobj \( (p = 0.030) \), and NegVOsubj \( (p = 0.017) \) at 2 weeks post thyroidectomy. This difference resolved within 3 to 6 months.

Predictive factors for early and late negative voice outcomes

Table 5 shows the following significant risk factors associated with early (2 week) and late (3 and 6 months) NegVO identified from univariate analysis of the thyroidectomy patients: extent of thyroidectomy; baseline characteristics; and histopathology data. Late NegVO were also correlated with the results of the voice outcomes measures at 2 weeks post thyroidectomy.

The extent of thyroidectomy \((TT > PT; p = 0.002)\) and size of resected thyroid gland \((114 vs 65 \text{ cm}^3; p = 0.048)\) were the only 2 significant predictors of early NegVO. Race \((p = 0.011)\) was the only significant predictor of late NegVO, with more NegVO in black patients around 3 to 6 months. The remainder of the baseline characteristics and histopathology data (age, BMI, sex, smoking status, voice professional status, inflammatory pathology, and malignant pathology) were not significantly associated with early or late NegVO.

Univariate analysis of the association of voice outcomes measures at the 2-week follow-up visit and late NegVO revealed that an abnormal VLS \((p < 0.001)\), a self-perceived “atypical” voice \((p = 0.010)\), and a change in...
VHI ≥20 from baseline (p = 0.003) at the 2-week follow-up visit were significant predictors of late NegVO, as seen in Table 5. A significant change from baseline in the DSI (≥2.49) and CAPE-V (≥11%) were not associated with late NegVO.

The ROC of changes in nominal voice outcomes measures (VHI, ΔVHI, ΔCAPE-V, and ΔDSI) at the 2-week follow-up visit and their association with late NegVO are seen in Figure 4. This analysis revealed that the change in VHI from baseline to 2 weeks (area under the curve 0.79; 95% CI, 0.64–0.95) and VHI score at the 2-week follow-up visit (area under the curve 0.75; 95% CI, 0.61–0.95) were most predictive of late NegVO.

Additional analysis, prompted by the results of this ROC evaluation, revealed that a VHI score of ≥18 at the 2-week follow-up visit was also a strong predictor of late NegVO, as seen in Table 5.

DISCUSSION
Timely recognition and intervention for postoperative dysphonia after thyroid surgery can help improve quality of life and possibly prevent maladaptive vocal behaviors. The ideal screening tool for functionally significant post-thyroidectomy voice disorders could be practically implemented perioperatively in the surgical outpatient clinic, be easily applied to individual patients, and be interpreted by surgeons. Such a tool would enable consistent clinical characterization of postoperative voice outcomes and initiation of timely referrals to speech pathology and otolaryngology as indicated by the tool’s assessment of dysphonia.40

Our group has developed a comprehensive longitudinal multidimensional voice outcomes classification tool that incorporates contributing variables from a prospectively validated set of commonly used voice outcomes measures.36,37,40,44-46,49-52 Significant changes in these measures are equally weighted in determining voice outcomes and used to characterize NegVO as either objective or subjective impairments. Differentiating NegVO patients into NegVOobj and NegVOsubj can assist in establishing a postoperative baseline and after progression of a specific impairment. This can be particularly useful in the early

Table 4. Voice Outcomes Compared over Time Using Generalized Estimating Equations in the Total Thyroidectomy and Partial Thyroidectomy Groups

<table>
<thead>
<tr>
<th>Voice outcomes</th>
<th>Total thyroidectomy, n (%)</th>
<th>Partial thyroidectomy, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2 wk (n = 53)</td>
<td>3 mo (n = 46)</td>
</tr>
<tr>
<td>NegVO</td>
<td>23 (43)</td>
<td>8 (17)</td>
</tr>
<tr>
<td>NegVO objective</td>
<td>11 (21)</td>
<td>5 (11)</td>
</tr>
<tr>
<td>NegVO subjective</td>
<td>12 (23)</td>
<td>3 (7)</td>
</tr>
<tr>
<td>NormVO</td>
<td>30 (57)</td>
<td>38 (83)</td>
</tr>
</tbody>
</table>

There were no negative voice outcomes in the non-neck surgery group.

*Significant.

†Outcomes with very few events for which generalized estimating equation results could not be reliably estimated.

NegVO, negative voice outcomes; NormVO, normal voice outcomes.

Figure 3. Incidence of voice outcomes for total thyroidectomy (TT) and partial thyroidectomy (PT) groups at each time point. Total thyroidectomy group with significantly more negative voice outcomes (NegVO)* (p = 0.002), objective NegVO* (p = 0.030), subjective NegVO* (p = 0.017) than PT at 2-week follow-up visit. Black bar, subjective negative voice outcomes; gray bar, objective negative voice outcomes.
postoperative period when dysphonia secondary to RLN dysfunction is suspected, and early vocal fold medialization by temporary augmentation could be beneficial.53,54

Results of this study support previous findings that postoperative voice dysfunction is common early after thyroid surgery. No NegVO were evident in the control NN group undergoing NN under general anesthesia. This latter finding, however, is contrary to earlier studies, which have reported 3% to 6% voice impairment after endotracheal intubation alone in NN patients.25,27,55 This difference could be attributed to the more stringent NegVO criteria (at least 2 abnormal outcomes measures) used in our study, and a smaller control group (n = 23 vs n = 100),27 which provided fewer opportunities for complications.

The extent of thyroidectomy was significantly associated with early (within 2 weeks of operation) NegVO in this study, with a higher rate of early NegVO in patients undergoing TT as compared with PT. A review of the literature identifies several studies that also report differences in voice alterations in the early postthyroidectomy period relative to the extent of thyroidectomy.26,56-58 The 2 most recent studies evaluated acoustic parameters and demonstrated a significant postoperative difference in voice measures between TT and PT patients, including fundamental frequency,26 maximum phonation time,56 shimmer,56,58 and noise-to-harmonic ratio.56,58 Interestingly, several of the observed differences in acoustic parameters relative to the extent of thyroidectomy in the early postoperative period were not concordant among these studies. In addition, 2 recent publications demonstrated no differences in acoustic parameters in the early postoperative period relative to the extent of thyroidectomy.10,59 It is difficult to reconcile the results of our analysis with these studies, given the discordant results identified in the published literature underscore the need for a multiparametric acoustic algorithm to quantify postoperative dysphonia, such as the recently published cepstral spectral index of dysphonia that was validated with data from the VOCAL trial.32

The greater incidence of NegVO in the TT than the PT groups at the 2-week postoperative follow-up visit

---

**Table 5. Factors Associated with Early (2 Weeks) and Late (3 to 6 Months) Negative Voice Outcomes Identified from Total Thyroidectomy and Partial Thyroidectomy Patients’ Baseline Characteristics, Pathology Data, and 2-Week Voice Outcomes Measure Results Based on Univariate Analysis**

<table>
<thead>
<tr>
<th>Factors associated with early negative voice outcomes</th>
<th>p Value</th>
<th>Odds ratio (95% CI)*</th>
<th>Negative predictive value, % (95% CI)*</th>
<th>Positive predictive value, % (95% CI)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extent of thyroidectomy</td>
<td>0.002*</td>
<td>5.9 (1.8–19.2)</td>
<td>88.6 (73.5–96.1)</td>
<td>43.4 (30.9–56.7)</td>
</tr>
<tr>
<td>Size of resected thyroid</td>
<td>0.048*</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Factors associated with late negative voice outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Race, black</td>
</tr>
</tbody>
</table>

2-week voice outcomes measures associated with late negative voice outcomes

| Abnormality on videolaryngostroboscopy              | <0.001* | 13.8 (3.3–57.1)      | 92.2 (82.6–97.0)                       | 53.8 (29.1–76.8)                      |
| Self-perceived atypical voice                       | 0.01*   | 6.0 (1.5–24.4)       | 93.6 (82.2–98.5)                       | 29.0 (15.9–46.8)                      |
| ΔVoice Handicap Index ≥20                           | 0.003*  | 7.3 (1.9–27.8)       | 92.7 (82.3–97.6)                       | 36.4 (19.6–57.1)                      |
| ΔVoice Handicap Index ≥18                           | 0.003*  | 9.8 (2.0–48.5)       | 95.6 (84.4–99.6)                       | 31.3 (17.8–48.7)                      |

*Significant. The remainder of the baseline characteristics and pathology data were not significantly associated with early or late negative voice outcomes.

1The size of resected thyroid gland in patients with early negative voice outcomes (median 114.2 cm³; interquartile range 62.0–217.0) was larger than the size of the resected thyroid gland in patients with early normal voice outcomes (median 65.0 cm³; interquartile range 28.8–165.8).

---

**Figure 4.** Receiver operating characteristics curve for change in Voice Handicap Index (VHI) score (area under the curve [AUC] 0.79; 95% CI, 0.64–0.95), VHI score (AUC 0.75; 95% CI, 0.61–0.95), change in Dysphonia Severity Index (DSI) (AUC 0.65; 95% CI, 0.47–0.82), and change in the overall severity of Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) (AUC 0.73; 95% CI, 0.56–0.89) at the 2-week follow-up as predictors of late (3 to 6 months) negative voice outcomes.
was true for both objectively verified (NegVOobj) and subjectively determined voice alterations (NegVOsubj). The former (NegVOobj) is intuitive, given that there are twice as many laryngeal nerves at risk in the TT compared with the PT group, and the mechanism for the latter (NegVOsubj) is not as clear. Because most voice impairments resolved within 6 months after surgery, the early difference in subjective voice impairments between groups could be attributed to more paratracheal and perithyroid dissection in the TT group leading to more inflammation, venous congestion, and laryngeal edema in the early postoperative setting. The finding that the size of the resected thyroid gland was also associated with early NegVO also supports this hypothesis.

Persistent differences between the 2 thyroidectomy groups in NegVO at 6 months (TT 16% vs PT 8%), although not statistically significant, could be associated with increased postoperative scarring and/or damage to smaller branches of the efferent laryngeal nerves in the TT group. A more invasive evaluation with laryngeal electromyography could be performed to evaluate such subtle changes, but this was beyond the scope of this study.

Of note, race was the only baseline characteristic found to be associated with late NegVO in this study. The exact mechanism for the increased NegVO in blacks remains unclear, possible explanations include a higher incidence of multinodular goiter in this population or racial differences in anatomic variants of the efferent laryngeal nerves. The former is unlikely, given that the presence of multinodular goiter has been demonstrated not to be a significant determinant of voice outcomes in VOCAL trial patients, and the latter has not been specifically studied in this patient population. Rather, the more likely explanation relates to cultural differences in the perception of the impact of voice disorders as demonstrated by Radowsky and colleagues and Yiu and colleagues.

The significant early changes in voice case history (atypical voice), VHI (change ≥18), and VLS (objective abnormality identified) were all found to be predictive of late NegVO after thyroidectomy. Additional analysis revealed that, in addition to these known markers of postoperative dysphonia, a VHI score of ≥18 at the 2-week follow-up visit was also strongly associated with late NegVO.

Based on these results, we suggest a clinical decision algorithm for the management of voice alterations in post-thyroidectomy patients, illustrated in Figure 5. By incorporating voice case history, VHI, and VLS into the standard preoperative and 2-week postoperative evaluation for patients undergoing thyroidectomy, predictors of late NegVO could be identified early in the postoperative period. This information could then be used to

![Clinical decision algorithm](image-url)
educate the patient about their likelihood of having a persistent voice disorder up to 6 months after thyroidectomy, which would trigger timely referral to voice-specialized professionals (eg, otolaryngologist and/or speech-language pathologist) for evaluation and consideration of vocal rehabilitation and/or vocal fold medialization.

The principal limitation of this study is that patient participation decreased between the 2-week and 6-month follow-up visits. This decline in participation was likely due to the significant time burden for multidimensional voice testing. Of note, only 2 of the patients who were lost to follow-up had a NegVOsubj at the 2-week visit, and the remainder of these patients had a normal voice outcomes at the last attended postoperative visit. Enrollment and retention of more patients into the VOCAL trial would have allowed for evaluation of a broader range of postoperative voice outcomes, and would have allowed for multivariate analysis of patient characteristics' influence on early and late NegVO.

A second key consideration is this study’s unique use of our voice outcomes classification tool, but the ability to compare these results to the literature outside of this study is limited. Given the internal validation of the voice outcomes measures and the use of the voice outcomes classification tool in this study, however, we have found that our results reflect a comprehensive perioperative voice evaluation. The value of our algorithm lies in its ability to include what we consider the cornerstone voice-specific parameters—videolaryngoscopic, acoustic, auditory perceptual, and quality of life indicators—as well as its tolerance of diverse inclusion criteria, such that a patient could be categorized or “binned” into 1 group, and have a slightly different voice impairment profile than another person in the same group. We have found that this classification tool has sufficient sensitivity to identify subtle subjective postoperative voice alterations, as well as clinically relevant specificity to exclude baseline voice alterations from nontraumatic general endotracheal anesthesia. The voice outcomes classification model appears to be an excellent clinical tool for conducting voice evaluation in surgical patients.

Another limitation to this study is that implementation of our comprehensive voice outcomes measures (ie, voice case history, VHI, CAPE-V, DSI, and VLS) would not be practical for most thyroid surgeons, given the need for speech-language pathologist interpretation of CAPE-V, as well as the equipment required for DSI. Interestingly, analysis of all outcomes measures revealed that the 2-week voice case history, VHI, and VLS were the most predictive voice outcomes measures post thyroidectomy for late adverse voice outcomes. These outcomes measures are readily available to, and easily applied by, most thyroid surgeons, and although the positive predictive value for late NegVO with these 3 measures was relatively low, all 3 measures retained excellent negative predictive value. As such, they would serve as practical screening tools to warrant additional voice evaluation by speech-language pathology and/or otolaryngology.

CONCLUSIONS
Since the inception of the VOCAL trial in 2004, our group has developed a comprehensive voice outcomes classification tool. In this specific analysis of the VOCAL Trial data, we have demonstrated that the extent of thyroidectomy impacts early voice outcomes. In addition, we have identified that black race, a perceived atypical voice, a Voice Handicap Index score >18, or an abnormality on videostroboscopic laryngeal examination are all predictive of persistent dysphonia. These factors can be practically evaluated in the early postoperative period, and at-risk patients can then be educated about their functional recovery, informed about their individual particular risk for late adverse voice outcomes, and referred to voice specialists in speech-language pathology and/or otolaryngology in a timely manner for early intervention to prevent development of maladaptive compensatory mechanisms, thereby improving the patient’s overall care and functional outcomes.

Author Contributions
Study conception and design: Solomon, Henry, Howard, Helou, Coppit, Shriver, Libutti, Shaha, Stojadinovic
Acquisition of data: Solomon, Helou, Coppit, Shriver, Libutti, Stojadinovic
Analysis and interpretation of data: Vicente, Solomon, Avital, Howard, Helou, Stojadinovic
Drafting of manuscript: Vicente, Solomon, Avital, Howard, Helou, Libutti, Shaha, Stojadinovic
Critical revision: Solomon, Avital, Henry, Helou, Coppit, Shriver, Libutti, Shaha, Stojadinovic

Acknowledgment: The authors owe a debt of gratitude for the concerted efforts and steadfast support of our research coordinator, Tiffany Felix, and the Henry M Jackson Foundation for the Advancement of Military Medicine.

REFERENCES


