Abstract

Objective. To assist otolaryngologists in counseling patients with hoarseness who would benefit from injection laryngoplasty on whether or not to perform the procedure in the office vs the operating room.

Data Sources. Cochrane library, CINAHL, PubMed, and EMBASE.

Review Methods. Systematic review using Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) reporting standards of English-language articles that compared voice outcomes for in-office and in the operating room injection laryngoplasty. Two independent investigators assessed study eligibility, rated the quality using Methodological Index for Non-Randomized Studies (MINORS), and abstracted data for comparative analysis.

Results. Of 689 initial studies, 4 observational, comparative studies met inclusion criteria, with follow-up of 2 weeks to 12 months postinjection. Laryngoplasty was most commonly performed for vocal fold immobility with varied injectable materials (micronized dermis, hyaluronic acid, and calcium hydroxyapatite). Follow-up ranged from 2 weeks to 12 months. Voice outcomes improved in all studies, with comparable improvement for patients injected in the office vs the operating room ($P = .42$ to $P = .88$). Meta-analysis of 3 studies showed no difference in Voice Handicap Index–10 voice outcomes by treatment setting (standardized mean difference $-0.11$, $P = .441$), with the 95% confidence interval ($-0.405$ to $0.176$), making it unlikely that anything larger than a small or trivial difference was missed.

Conclusion. Our systematic review makes it unlikely that meaningful clinical differences exist in postprocedure voice outcomes for injection laryngoplasty in the office vs the operating room.

Keywords

injection laryngoplasty, vocal fold immobility, voice, dysphonia, hoarseness, meta-analysis
Materials and Methods

Literature Search Strategy

The literature review was performed according to an a priori protocol that adhered to current standards for the conduct and reporting of systematic reviews, as suggested by the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) Statement. An electronic database search of PubMed, CINAHL, EMBASE, and Cochrane Library was performed for articles published between November 1978 and November 2017. The search was designed to answer the question: for patients undergoing IL, is there a difference in voice outcomes when comparing local to general anesthesia? An information specialist was consulted to define appropriate search terms, which included injection laryngoplasty, vocal fold injection, vocal cord injection, voice, and office. A keyword-based search was performed using all possible combinations of search terms. Results were limited to English-language articles. References of all included studies were reviewed to ensure no related studies were overlooked. Institutional review board approval was not required due to the nature of the study.

Selection Criteria

The review aimed to identify studies that compared voice outcomes for IL performed under local vs general anesthesia. Two investigators, working independently, reviewed study abstracts and excluded studies that were not published as full-length articles, only included patients younger than 18 years, included nonhuman subjects, did not include voice outcomes, or were limited to one type of anesthesia, without a comparison group. Full-length articles were then obtained for remaining studies and were excluded if they did not quantify voice outcomes or contained insufficient voice data for analysis.

Data Extraction and Risk-of-Bias Assessment

Two reviewers independently abstracted data from included articles using a standardized form. Descriptive results

Figure 1. Study identification flow diagram.

*Refers to the exclusion criteria (1) in the figure, listed under “Exclusion Criteria.”

*Refers to the exclusion criteria (2) in the figure, listed under “Exclusion Criteria.”
included country of study, study design, number of patients studied, number of patients included in voice outcomes, patient age and sex, number of general and local anesthesia, most common indication for IL, tools used to measure voice outcomes, and length of follow-up. Voice outcomes were compared using the pre- and postinjection scores for local and general anesthesia groups. The risk of bias for each study was assessed by 2 investigators independently using the Methodological Index for Non-Randomized Studies (MINORS).5

**Statistical Analysis and Outcome Measurements**

Study results were pooled using a random-effects model of meta-analysis, which assumes the true effect could vary from study to study.6 The primary outcome was change in the Voice Handicap Index–10 (VHI-10) before vs after injection laryngoscopy for the 2 study groups (general anesthesia vs local anesthesia), expressed as the standardized mean difference (positive values favor general anesthesia). The meta-analysis was performed using sample size, mean change score, and standard deviation of the change score for each group. Two studies7,8 provided standard deviations, but for 1 study,9 the standard deviation was calculated from the confidence interval.10 A forest plot was prepared for the pooled analysis, showing study results, 95% confidence intervals, and the pooled effect size. Heterogeneity was assessed and reported using the I² statistic, where 25% is low, 50% is moderate, and 75% is high heterogeneity among studies.11 Statistical significance was assessed with a type I error probability (P value) threshold of .05.

**Results**

**Literature Review and Risk-of-Bias Assessment**

The initial search identified 689 studies, of which 309 were unique and not duplicates (Figure 1). Four articles met inclusion criteria.7,9,12 Study characteristics and demographics are described in Table 1. A total of 389 participants were enrolled, of which 268 completed voice outcome surveys. Three studies used retrospective cohort design, and there was 1 prospective cohort. Risk-of-bias assessment for each study using the MINORS instrument is listed in Table 2. According to the authors of the MINORS instrument, comparative, nonrandomized studies that were rated as “excellent” had a mean global score of 19.8.5 For our review, global scores ranged from 17 to 18 out of 24, indicating a low degree of bias. All studies lost points for lack of blinding and failure to calculate study design prospectively, common weaknesses of observational studies. However, the studies did include adequate control groups, consecutive patients, appropriate length of follow-up, and suitable outcome measures, which strengthens confidence in their conclusions.

**Study Characteristics and Voice Outcomes**

Three studies measured voice outcomes using the VHI-10, and 1 study used the Voice-Related Quality of Life (V-RQOL). The most common indication for IL was unilateral vocal fold weakness. Injectable materials included calcium hydroxyapatite, hyaluronic acid, and micronized alloderm. Voice outcomes for each study are listed in Table 3. For all 4 studies, there was a statistically significant improvement in voice following IL, for both local and general anesthesia groups. However, no study reported a significant difference in voice scores when comparing local vs general anesthesia. The difference in preoperative (baseline) scores between the 2 groups was not statistically significant for any of the studies, although differences cannot be excluded due to the small sample sizes.

**Meta-Analysis**

A pooled analysis was obtained for the 3 studies using the VHI-10 tool to assess voice outcomes (Figure 2). The pooled standardized mean difference (SMD) of –0.0114 favors the local anesthesia group, but this result was not statistically significant (P = .441). The I² is 0%, indicating a

### Table 1. General Characteristics and Demographic Data of Included Studies.

<table>
<thead>
<tr>
<th>Study, Year</th>
<th>Study Design</th>
<th>Total No.</th>
<th>No. Included in Voice Outcomes</th>
<th>Mean Age, y</th>
<th>Sex (M/F), No.</th>
<th>Anesthesia (LA/GA), No.</th>
<th>Most Common Indication</th>
<th>Most Common Injectate</th>
<th>Outcome Measures</th>
<th>Length of Follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chandran et al9</td>
<td>Retrospective cohort</td>
<td>75</td>
<td>73</td>
<td>64.6</td>
<td>36/39</td>
<td>41/34</td>
<td>NR</td>
<td>Hyaluronic acid</td>
<td>VHI-10, GRBAS</td>
<td>3 wk</td>
</tr>
<tr>
<td>Zelenik et al8</td>
<td>Prospective cohort</td>
<td>31</td>
<td>31</td>
<td>60.9</td>
<td>12/19</td>
<td>17/14</td>
<td>UVFP</td>
<td>Calcium hydroxyapatite</td>
<td>VHI-10, MPT</td>
<td>12 mo</td>
</tr>
<tr>
<td>Mathison et al12</td>
<td>Retrospective cohort</td>
<td>141</td>
<td>78</td>
<td>61.3</td>
<td>83/58</td>
<td>105/61</td>
<td>UVFP</td>
<td>Micronized alloderm</td>
<td>V-RQOL</td>
<td>2-8 wk</td>
</tr>
<tr>
<td>Bove et al7</td>
<td>Retrospective cohort</td>
<td>142</td>
<td>86</td>
<td>60.7</td>
<td>103/39</td>
<td>45/97</td>
<td>UVFP</td>
<td>Calcium hydroxyapatite</td>
<td>VHI-10</td>
<td>2 mo</td>
</tr>
</tbody>
</table>

Abbreviations: GA, general anesthesia; GRBAS, grade, roughness, breathiness, asthenia, strain scale; LA, local anesthesia; MPT, maximum phonation time; NR, not reported; UVFP, unilateral vocal fold paralysis; VHI-10, Voice Handicap Index–10; V-RQOL, voice-related quality of life survey.
high degree of homogeneity among included studies. The upper limit of the pooled 95% confidence interval (SMD of 0.176) is consistent with no difference or possibly a trivial difference, as interpreted using Cohen’s effect size. The lower limit of the pooled 95% confidence interval (SMD of −0.405) is consistent with a small benefit favoring local anesthesia. Although this means we cannot exclude the possibility of some benefit for local anesthesia over general

Table 2. Assessment of Risk of Bias Using the Methodological Index for Non-Randomized Studies (MINORS) Instrument.

<table>
<thead>
<tr>
<th>MINORS Criteria</th>
<th>Chandran et al⁹</th>
<th>Zelenik et al⁸</th>
<th>Mathison et al¹²</th>
<th>Bove et al⁷</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. A clearly stated aim: the question addressed should be precise and relevant in the light of available literature.</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2. Inclusion of consecutive patients: all patients potentially fit for inclusion (satisfying the criteria for inclusion) have been included in the study during the study period (no exclusion or details about the reasons for exclusion).</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3. Prospective collection of data: data were collected according to a protocol established before the beginning of the study.</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4. Endpoints appropriate to the aim of the study: unambiguous explanation of the criteria used to evaluate the main outcome that should be in accordance with the question addressed by the study. Also, the endpoints should be assessed on an intention-to-treat basis.</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>5. Unbiased assessment of the study endpoint: blind evaluation of objective endpoints and double-blind evaluation of subjective endpoints. Otherwise, the reasons for not blinding should be stated.</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6. Follow-up period appropriate to the aim of the study: the follow-up should be sufficiently long to allow the assessment of the main endpoint and possible adverse events.</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>7. Loss to follow-up less than 5%: all patients should be included in the follow-up. Otherwise, the proportion lost to follow-up should not exceed the proportion experiencing the major endpoint.</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>8. Prospective calculation of the study size: information of the size of detectable difference of interest with a calculation of 95% confidence interval, according to the expected incidence of the outcome event, and information about the level for statistical significance and estimates of power when comparing the outcomes.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9. An adequate control group: having a gold standard diagnostic test or therapeutic intervention recognized as the optimal intervention according to the available published data.</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>10. Contemporary groups: control and studied group should be managed during the same time period (no historical comparison).</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>11. Baseline equivalence of groups: the groups should be similar regarding the criteria other than the studied endpoints. Absence of confounding factors that could bias the interpretation of the results.</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>12. Adequate statistical analyses: whether the statistics were in accordance with the type of study with calculation of confidence intervals or relative risk.</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>18/24</td>
<td>18/24</td>
<td>18/24</td>
<td>17/24</td>
</tr>
</tbody>
</table>

anesthesia, we can conclude with good confidence that there is no benefit to general anesthesia vs local anesthesia based on these 3 studies.

### Discussion

This study had the primary objective of assessing voice outcomes following IL, comparing results of procedures performed under local vs general anesthesia. While several studies have reviewed the indications, complications, and approaches for awake and asleep IL, this is the first study to assess voice outcomes.

The estimated point prevalence of voice disorders in the United States is 20 million (0.98%), resulting in over $5 billion in annual direct costs. This figure includes medication, procedure, and medical encounter claims but does not account for lost productivity. Any effort to improve quality of life and reduce costs requires accurate assessment of treatment outcomes.

A variety of methods are used to assess outcomes following surgical treatment of voice disorders, including laryngoscopy; auditory-perceptual tools such as the Grade, Roughness, Breathiness, Asthenia, Strain scale; and patient-reported outcome (PRO) measures. Although there are a number of objective methods used to evaluate the success of IL, no objective test replaces the importance of a patient’s perspective. The use of patient-reported quality-of-life (QOL) measures has become an essential part of clinical practice. The US Food and Drug Administration (FDA) recommends the use of QOL measures for the development of pharmacological products and devices, and the National Institutes of Health (NIH) has identified PRO/QOL measures as critical to drive progress in biomedical research. In addition, preapproval and reimbursement by insurance companies is increasingly linked to standardized outcome measures. Although there is a variable degree of methodological rigor in the development of these measures, their use in all medical specialties has become widespread. PRO measures serve a key role for evaluating treatment effectiveness in voice-related diseases.

A number of voice outcome assessment tools have been used to measure the subjective change in voice quality after IL. Common symptom-specific surveys include the VHI-10, V-RQOL, Voice Performance Questionnaire (VPQ), and the Voice Outcome Survey (VOS). All 4 included studies demonstrated an improvement in voice outcomes following IL, with no significant difference between general and local anesthesia. For the 3 studies included in meta-analysis, the pooled standardized mean difference (SMD) of −0.114 indicates a trivial difference and was not statistically significant. While the pooled results suggest that there is no benefit of general anesthesia, we cannot exclude the possibility of a small benefit for local anesthesia. It is important to consider the possible factors that may skew these results. While there was no statistically significant difference in preoperative voice outcome scores, the small sample size allows for the possibility of undetected differences in baseline disease severity. That is, patients with more severe vocal fold dysfunction may be

<table>
<thead>
<tr>
<th>Study, Year</th>
<th>Subjective Voice Assessment Tool</th>
<th>Local Anesthesia</th>
<th>General Anesthesia</th>
<th>Best Score (SD; CI)</th>
<th>Preinjection, Mean (SD)</th>
<th>Postinjection, Mean (SD)</th>
<th>Change, Mean Best Score (SD; CI)</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chandran, et al.</td>
<td>VHI-10</td>
<td>0 of 40</td>
<td>27.1 (8.95)</td>
<td>19.9 (10.8)</td>
<td>7.27 (10.1; 4.5-10.0)</td>
<td>.388</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zelenik, et al.</td>
<td>VHI-10</td>
<td>0 of 40</td>
<td>27.1 (8.95)</td>
<td>19.9 (10.8)</td>
<td>7.27 (10.1; 4.5-10.0)</td>
<td>.388</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mathison, et al.</td>
<td>V-RQOL</td>
<td>0 of 40</td>
<td>27.1 (8.95)</td>
<td>19.9 (10.8)</td>
<td>7.27 (10.1; 4.5-10.0)</td>
<td>.388</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bove, et al.</td>
<td>VHI-10</td>
<td>0 of 40</td>
<td>27.1 (8.95)</td>
<td>19.9 (10.8)</td>
<td>7.27 (10.1; 4.5-10.0)</td>
<td>.388</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: CI, 95% confidence interval; GA, general anesthesia; LA, local anesthesia; NR, not reported; SD, standard deviation; VHI-10, Voice Handicap Index–10.
preferentially treated under general anesthesia. Prospective, randomized controlled trials are necessary to minimize the risk of bias in patient selection. Although some uncertainty in the results remains, the pooled analysis improves our confidence over the individual studies, and we can conclude that local anesthesia is at least comparable to general anesthesia in terms of effect on VHI-10.

When counseling a patient regarding the risks, benefits, and alternatives of performing the IL under general or local anesthesia, a number of patient-related factors must be discussed. However, given the intent of the procedure to improve perceived voice quality, the possible difference in voice outcomes between the 2 techniques must be an essential part of the discussion. This conclusion enables otolaryngologists to use the approach that is most appropriate for the individual patient, with assurance that anesthetic type will not negatively affect voice outcomes. Local anesthesia offers several advantages: avoidance of general anesthesia, lower cost, and immediate assessment of voice during the procedure. However, many patients cannot tolerate awake procedures, due to anxiety, poor tolerance, unfavorable anatomy, or strong gag reflex. In addition, general anesthesia facilitates additional grafting techniques.

Two studies reviewed complications. Chandran et al found a lower rate of complications with local anesthesia (LA) (4.9%) than general anesthesia (GA) (14.7%), while Mathison et al found higher complication rates with LA (19.1%) compared to GA (6.6%). Complications of local anesthesia tend to arise from inadequate visualization, inadvertent vocal fold movement, or excessive bleeding. As a result, the most common reported complications for awake IL were abortion of the procedure and inadvertent injection of material into the superficial lamina propria.

Although the results of this review suggest comparable voice outcomes for LA and GA, there are several limitations to consider. First, only 4 studies were included in the review and only 3 included in pooled analysis. The small number of studies that met inclusion criteria reflects the focused question and rigorous methodology of this review but also the paucity of quality studies investigating this topic. While a number of published articles report voice outcomes for IL, few compare local to general anesthesia. Nonetheless, the homogeneity of the studies improves confidence in the results of meta-analysis. The results of this review should highlight the need for more randomized, prospective studies to better determine the difference in voice outcomes when IL is performed under general or local anesthesia. Second, although procedures were largely performed by 1 surgeon, the techniques and injection materials varied, which could potentially influence results. Finally, all 4 studies used a cohort design, 3 of which were performed retrospectively. None of the studies randomly assigned patients to treatment groups, and analysis of outcomes was not blinded. These weaknesses are reflected in the risk of bias assessment (MINORS) scores, where points were lost for lack of blinding and retrospective design. Further investigation of this topic with randomized controlled trials would strengthen the evidence supporting comparable voice outcomes.

Conclusion

Our systematic review makes it unlikely that large differences exist in postprocedure voice outcomes for IL in the office vs the operating room, but the limited evidence has inadequate statistical power to exclude smaller differences. While each study showed an improvement in subjective voice scores after IL, there was no significant difference when comparing local to general anesthesia. These results can help facilitate discussion with patients when selecting candidates for local vs general anesthesia for IL. As in-office laryngeal procedures become increasingly common, more rigorous studies into this topic are necessary to improve confidence that voice outcomes are comparable for the 2 techniques.

Author Contributions

Daniel P. Ballard, conception and design; acquisition, analysis, and interpretation of data; drafting and revision of the manuscript; final approval; Jason Abramowitz, conception and design; acquisition, analysis and interpretation of data; revision of the manuscript; final approval; Daniel C. Sukato, acquisition, analysis, and
interpretation of data; revision of the manuscript; final approval; Boris Bentsianov, conception and design; revision of the manuscript; final approval; Richard Rosenfeld, conception and design; analysis and interpretation of data; revision of the manuscript; final approval.

Disclosures

Competing interests: None.
Sponsorships: None.
Funding source: None.

References