The Laryngoscope

Long-Term Outcomes in Unilateral Vocal Fold Paralysis Patients

Conor W. McLaughlin, MD; Brian Swendseid, MD; Mark S. Courey, MD;
Sarah Schneider, MS, CCC-SLP; Jackie L. Gartner-Schmidt, PhD; Katherine C. Yung, MD

Objectives/Hypothesis: At presentation, unilateral vocal fold paralysis (UVFP) patients have different treatment options, including conservative management (CM), injection laryngoplasty (IL) with a temporary agent, or permanent medialization (PM). This study evaluates long-term outcomes for UVFP patients relative to intervention.

Study Design: Retrospective chart review.

Methods: A retrospective chart review was performed of UVFP patients who presented to the University of California San Francisco Voice and Swallowing Center. Videolaryngostroboscopy examinations were reviewed. Maximum glottic closure was quantified with the normalized glottic gap area (NGGA). Perceptual voice analysis was performed using the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) at corresponding time points.

Results: Fifty-three patients met inclusion criteria. Six underwent CM only, 20 went on to require PM, 19 underwent IL only, and eight underwent IL and subsequent PM. NGGA at presentation was similar among groups; however, the CM group was noted to have more favorable CAPE-V scores for Breathiness ($P = .007$) and Loudness ($P = .018$). All groups had similar NGGA and CAPE-V scores at last follow-up. When compared to pooled data for patients who underwent PM, the IL group was noted to have similar NGGA and CAPE-V scores at presentation. Although improvements in both groups were noted following intervention, both groups appeared similar at last follow-up with the exception of Roughness, for which the IL group retained a slightly improved outcome (13.3 vs. 18.3, $P = .03$).

Conclusions: At presentation, UVFP patients have similar NGGA. This finding suggests that treatment recommendations cannot be made on the size of the glottic gap alone. Furthermore, in many patients, IL results in long-term benefit with glottic closure and CAPE-V scores equivalent to that obtained with PM.

Key Words: Vocal fold paralysis, injection laryngoplasty, type I thyroplasty.

Level of Evidence: 4.

***Laryngoscope*, 128:430–436, 2018**

INTRODUCTION

Injury to the vagus nerve or to the recurrent laryngeal nerve branch may result in loss of functional vocal fold mobility. In such patients with unilateral vocal fold paralysis (UVFP), the inability of the affected vocal fold to properly adduct and contact the functioning contralateral vocal fold leads to varying degrees of glottic incompetence. Consequences of UVFP often manifest as breathy dysphonia, with some patients additionally reporting symptoms of mild dysphagia. If the vagus nerve is injured, dysphagia may be more severe due to the impairment in both motor and sensory function.

For patients who present with UVFP, the commonly held treatment paradigm usually consists of one of three initial treatment strategies: 1) observation for spontaneous return of function or compensation, 2) voice therapy to maximize vocal efficiency and/or to facilitate compensation, and 3) injection laryngoplasty (IL) with a temporary agent designed to improve the position of the paralyzed vocal fold to facilitate glottic competence and improve voice while compensation or recovery occur.

Although histologic studies in animal models have demonstrated that commonly used temporary injection materials are usually absorbed by 6 months following injection, some studies have suggested that patients who undergo IL with a temporary agent have sustained clinical benefit that persists beyond this often-cited timeframe.[1–4] Specifically, two recent studies demonstrated that patients with UVFP who undergo IL with a temporary agent were less likely to subsequently undergo a permanent medialization (PM) procedure when

From the Department of Otolaryngology–Head and Neck Surgery (C.W.M., S.S., K.C.Y.), and Department of General Surgery (B.S.), University of California San Francisco, San Francisco, California; Department of Otolaryngology–Head and Neck Surgery (M.S.C.), Icahn School of Medicine at Mount Sinai, New York, New York; and the Department of Otolaryngology (J.L.G.-S.), University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, U.S.A.

Editor's Note: This Manuscript was accepted for publication on August 14, 2017.

Presented as a poster presentation at the American Laryngological Association Annual Meeting at the Combined Otolaryngology Spring Meetings (COSM), Chicago, Illinois, U.S.A., May 18–22, 2016.

This work was performed at the Voice and Swallowing Center, Department of Otolaryngology–Head and Neck Surgery, University of California San Francisco, San Francisco, California, U.S.A.

This work was supported by the National Center for Advancing Translational Sciences, National Institutes of Health, through UCSF-CTSI Grant Number UL1 TR000004. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

The authors have no other funding, financial relationships, or conflicts of interest to disclose.

Send correspondence to Katherine C. Yung, MD, UCSF Voice and Swallowing Center, Department of Otolaryngology–Head and Neck Surgery, University of California San Francisco, 2330 Post Street, 5th Floor, San Francisco, CA 94115. E-mail: katherine.yung@ucsf.edu

DOI: 10.1002/lary.26900

compared to those patients who opted for conservative management (CM) alone.[2,3] Furthermore, Prendes et al. demonstrated that this lower rate of permanent medialization in patients who underwent early injection with a temporary agent correlated with improved objective measures of glottic competence and perceptual voice analyses.[4] The mechanism by which an injection with a resorbable agent improves glottic competence beyond the accepted timeframe for material absorption and decreases rates of subsequent permanent medialization has not been fully elucidated.

Contrary to these findings, a recent retrospective cohort study by Francis et al. found that early IL did not affect the odds of ultimately needing open framework surgery in patients with UVFP who presented within 9 months of symptoms onset.[5] Despite conflicting evidence on the role of early IL and the subsequent need for PM, still less is known regarding how patients who receive early IL with a temporary agent compare to those individuals who undergo PM, either as initial treatment or as subsequent therapy. The purpose of this study, therefore, was to compare objective outcomes of glottic competence among those patients who underwent IL alone versus those patients who underwent CM or PM for UVFP.

MATERIALS AND METHODS

A retrospective chart review was performed on all patients presenting to the University of California San Francisco (UCSF) Voice and Swallowing Center with a diagnosis of UVFP between July 1, 2004 and July 1, 2014. Approval from the UCSF Committee on Human Research was obtained. Basic patient characteristics were recorded and included the following: age, gender, race, etiology of recurrent laryngeal nerve (RLN) or vagus nerve injury, date of injury or date of symptom onset, date of presentation to the UCSF Voice and Swallowing Center, prior treatments, type of initial and subsequent interventions, and time from beginning of treatment to the most recent follow-up examination.

Patients were stratified into one of four possible treatment groups based on intervention type: 1) patients who received CM alone, which consisted of observation for return of vocal fold mobility and voice therapy; 2) patients who began with CM but went on to require a PM procedure (CM + PM); 3) patients who underwent initial IL with a temporary agent alone; and 4) patients who initially underwent IL with a temporary agent but eventually required a permanent medialization procedure (IL + PM). Temporary injectable materials included Radiesse Voice Gel (Merz Aesthetics, Raleigh, NC), Cymetra (LifeCell, Branchburg, NJ), and Restylane (Galderma Laboratories, Fort Worth, TX). Permanent medialization was performed via type I thyroplasty with or without arytenoid adduction. There were certain instances in which patients declined open framework surgery or were not felt to be candidates for an open procedure (i.e., significant medical comorbidities). In these cases, patients were offered a longer acting vocal fold injection with calcium hydroxyapatite (CaHA) (Radiesse Voice; Merz Aesthetics) instead of open framework surgery. For the purposes of the present study, these patients were considered to have undergone a PM procedure. Although the duration of benefit of CaHA injection is finite, the treatment objective is long-term in contrast to the reabsorbable materials noted above, with one recent study demonstrating a mean duration of benefit after CaHA injection of 18.6 months (range, 8–36 months).[6–8]

Voice samples and videolaryngostoboscopic recordings were obtained in one of three clinical exam rooms at the UCSF Voice and Swallowing Center. Although efforts were made to minimize ambient noise during examination, these clinic rooms were not rendered soundproof at the time of data acquisition. Patients were asked to hold the microphone at a comfortable distance from their mouths during phonatory tasks; however, it should be noted that this distance was not standardized across patient encounters. Voice recordings were obtained on a nondirectional electret condenser microphone (Olympus ME-15; Olympus Medical, Center Valley, PA). Each patient was asked to read a standard passage and produce sustained vowel sounds at a conversational pitch and loudness. Videolaryngostroboscopic recordings were obtained with a 70° rigid telescope (or flexible rhinolaryngoscope when necessary) connected to a high-definition narrow-band imaging camera head (CH-S190-XZ; Olympus Medical). Illumination was obtained with a stroboscopic xenon light source, and videos were processed with a narrow-band video processor (Evis Exera III CLV-190 and CV-190; Olympus Medical). Each patient was asked to proceed through multiple phonatory cycles of sustained /eee/ sound at various pitches. Both the voice recordings and videolaryngostroboscopic exams were archived on a secured video management server (Vault Stream; Image Stream Medical, Littleton, MA).

Digital voice samples and videolaryngostroboscopy recordings for each patient were retrieved at various time points throughout treatment from these archived files. Time points of interest included initial presentation to clinic, pre- and postprocedure follow-ups (where applicable for CM + PM, IL, and IL + PM groups), and last follow-up. Patients were excluded from analysis if they did not have at least 6 months of follow-up from presentation (if CM group) or from last intervention (if CM + PM, IL, or IL + PM groups). In addition, we excluded patients with a history of laryngeal trauma, history of laryngeal cancer or irradiation to the head and neck, prior treatment at outside facility, posterior glottic stenosis, and prior RLN resection of spasmodic dysphonia.

Voice samples were presented to two independent speech and language pathologists (SLPs) at all time points and were graded according to the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) criteria.[9,10] All voice samples were randomly presented and were in a deidentified format so that no inference could be made regarding the patient identity, intervention status, or time point during treatment. To obtain intrarater reliability data, 10% of the voice recordings were selected using a random number generator for repeat analysis by the graders. Again, grading was carried out in a blinded manner as detailed above.

Videolaryngostroboscopic recordings at time points of interest were reviewed by the primary authors (C.W.M. and B.S.). The point of maximum glottic closure was assessed while the patient produced the /eee/ sound at modal pitch and conversational loudness. For patients in whom modal pitch was difficult to assess, or for those patients unable to trigger the strobe, best estimates of the maximum glottic closure were made following a complete video review noting the point of maximum vocal fold adduction. Still images were captured at the period of maximum glottic closure and were analyzed using the image software ImageJ (National Institutes of Health, Bethesda, MD).[11] The normalized glottic gap area (NGGA) was calculated using the method outlined by Omori et al.[12] The glottic gap was traced and measured in square pixels. The length of the membranous vocal fold was measured from the vocal process to the anterior commissure and was expressed in linear pixels. The normalized glottic gap was calculated using the following

TABLE I.
Comparison of Demographic, Etiology, and Chronological Characteristics for Treatment Groups.

	CM	CM + PM	IL	IL + PM	*P* Value
No. of patients	6	20	19	8	
Age (yr)	60.2	70.7	65	52.5	.047*
Gender (% male)	16.7	40	26.3	37.5	.650
Etiology (%)					
Surgery	50	40	84.2	100	.003*
Tumor	16.7	10	0	0	.300
Neurologic event	0	15	0	0	.150
Idiopathic	33.3	30	10.5	0	.160
Other	0	5	5.3	0	.860
Side of paralysis (% left)	100	60	42.1	50	.091
Median time to presentation (mo)	28.1	14.2	9.6	3.6	.018*
Median time to last follow-up (mo)	28.0	23.7	19.4	21.5	.720

*Effects with $P < .05$.
Comparisons between groups performed with Kruskal-Wallis test or χ^2 test, where applicable.
CM = conservative management only; CM + PM = conservative management followed by permanent medialization; IL = temporary injection laryngoplasty only; IL + PM = temporary injection laryngoplasty followed by permanent medialization.

equation: NGGA = glottic gap (pixels)^2 / [membranous vocal fold length (pixels)]^2 × 100.

Statistical analysis was performed using Stata statistical software (StataCorp, College Station, TX). To determine intra-rater and inter-rater reliability, Pearson correlation coefficients were calculated. For continuous variables and to compare difference in means between groups, one-way analysis of variance or χ^2 analyses were used for parametric data. When normality was not demonstrated, Kruskall-Wallis tests were used to compare differences in medians among groups. The level of significance was defined as $P < .05$.

RESULTS

A review of the UCSF Voice and Swallowing Center records identified 269 patients with a diagnosis of UVFP from 2004 to 2014. Of these patients, a total of 53 met all inclusion criteria and were stratified into the following treatment groups: CM (n = 6), CM + PM (n = 20), IL (n = 19), and IL + PM (n = 8). Reasons for patient exclusion included insufficient follow up (n = 136), inability to access videostroboscopic data or voice recordings (n = 29), history of radiation to the neck (n = 21), prior treatment at outside facility (n = 17), spontaneous recovery of function (n = 11), and resection of the RLN for spasmodic dysphonia (n = 2). Demographic characteristics, etiology, and chronological characteristics, including time to presentation and time to last follow-up, for each of the four groups are listed in Table I. Median time to last follow-up was 22.3 months for all patients (range, 19.4–28 months). There was a statistically significant difference in age and median time to presentation among groups, with the IL + PM group noted to be younger (average age 52.5 years, $P = .047$) and with shorter time to initial presentation (3.6 months, $P = .018$). In addition, surgery was the sole etiology for UVFP in the IL + PM group, which reached statistical significance (8/8, $P = .003$).

Intra- and inter-rater reliability data were calculated for our independent SLP graders of CAPE-V parameters. Intrarater reliability was assessed by way of Pearson correlation coefficients (r) based on a random re-presentation of 10% of voice samples to each SLP in a blinded format. Correlation coefficient values are shown in Table II. Our results indicated strong intrarater reliability among all CAPE-V characteristics except for Strain for SLP 2. Inter-rater reliability was similarly calculated using the Pearson correlation method and is also shown in Table II. Inter-rater reliability was noted to be strong among all CAPE-V characteristics, as indicated by $r > 0.5$, except for Strain and Roughness.

NGGA and average CAPE-V scores at presentation for each of the four groups are summarized in Table III. There was no statistically significant difference among NGGA between groups at presentation. In addition, average CAPE-V scores were similar among all groups at presentation with the exception of the CM group, who appeared to have more favorable CAPE-V scores for Breathiness (10.7, $P = .007$) and Loudness (2.8, $P = .018$) compared with the other groups. Table IV shows the NGGA and average CAPE-V scores at last follow-up for each of the four groups. Although all groups

TABLE II.
Reliability Data for Consensus Auditory-Perceptual Evaluation of Voice Analyses.

Characteristic	Intrarater (*r*) SLP 1	Intrarater (*r*) SLP 2	Inter-rater (*r*)
Overall severity	0.96	0.95	0.85
Roughness	0.59	0.79	0.45
Breathiness	0.94	0.94	0.78
Strain	0.82	−0.19*	0.30
Pitch	0.91	0.70	0.57
Loudness	0.69	0.84	0.71

Intrarater and inter-rater reliability was calculated by Pearson correlation coefficient (*r*).
*Effects with $P < .05$.
SLP = speech and language pathologist.

TABLE III.
Normalized Glottic Gap Area and Average CAPE-V Scores at Presentation Stratified by Treatment Group.

Variable	CM (n = 6)	CM + PM (n = 20)	IL (n = 19)	IL + PM (n = 8)	*P* Value
Normalized glottic gap area	1.79	5.72	4.66	3.57	.120
CAPE-V characteristic					
Overall severity	24.9	59.4	53.4	60.6	.070
Roughness	16.9	20.2	23.9	25.7	.410
Breathiness	10.7	49.8	42.7	55.8	.007*
Strain	8.0	17.8	20.4	14.4	.470
Pitch	11.8	27.1	27.3	19.8	.390
Loudness	2.8	35.7	29.2	39.9	.018*

Comparisons between groups performed with the Kruskal-Wallis test or χ^2 test as applicable.
*Effects with $P < .05$.
CAPE-V = Consensus Auditory-Perceptual Evaluation of Voice; CM = conservative management only; CM + PM = conservative management followed by permanent medialization; IL = temporary injection laryngoplasty only; IL + PM = temporary injection laryngoplasty followed by permanent medialization.

demonstrated an improvement in average NGGA and CAPE-V characteristic scores, there were no statistically significant differences among NGGA and average CAPE-V scores at last follow-up between groups.

We also performed comparisons between the IL group and those groups that went on to receive a PM procedure (CM + PM and IL + PM). Patients in the latter groups were pooled together and collectively referred to as "Any PM." Specifically, as noted in Table V, we compared NGGA and average CAPE-V scores at presentation and last follow-up for these two groups of patients. We found no statistically significant differences in NGGA at presentation between both groups. At last follow-up, although there were improvements across all parameters for both groups, there remained no statistically significant difference between average NGGA. In addition, average CAPE-V scores were similar at last follow-up among all groups. The exception was Roughness at last follow-up, with the IL group demonstrating slightly improved average CAPE-V scores when compared to the Any PM group (13.3 vs. 18.3, $P = .03$).

DISCUSSION

In patients with a diagnosis of UVFP with unknown prognosis for recovery of function, our group and others have previously shown that early IL with a temporary agent is associated with lower rates of subsequent PM procedures.[2,3] Furthermore, a recent study has suggested that these lower rates of PM in patients who undergo early IL with a temporary agent are associated with improved measures of glottic competence as judged by videolaryngostroboscopic assessments of glottic closure in addition to improved perceptual evaluations of voice.[4] Contrary to these findings, others have found that early IL has no effect on the ultimate rate of PM in patients who present early (<9 months) from symptom onset.[5] Although there remains the question as to whether early IL can obviate the need for subsequent PM in certain patient cohorts, less is known regarding long-term outcome measures among those patients who undergo IL with a temporary agent only and those who receive a PM procedure. Herein, we describe long-term outcomes for patients who received upfront IL with a temporary agent for a diagnosis of UVFP as compared with those patients who received conservative management alone or who subsequently required PM.

Our patient groups were similar with regard to gender distribution, sidedness of injury, and time to last follow-up. Of note, patients in the IL + PM group were noted to have a younger average age at presentation (52.5 years, $P = .047$) and a shorter median time to

TABLE IV.
Normalized Glottic Gap Area and Average CAPE-V Scores at Last Follow-up Stratified by Treatment Group.

Variable	CM (n = 6)	CM + PM (n = 20)	IL (n = 19)	IL + PM (n = 8)	*P* Value
Normalized glottic gap area	0.97	1.06	1.21	1.43	.920
CAPE-V characteristic					
Overall severity	29.5	32.9	25.4	23.5	.310
Roughness	17.2	19.1	13.3	15.0	.140
Breathiness	11.8	12.2	12.5	16.8	.460
Strain	10.9	14.4	14.0	6.1	.630
Pitch	16.0	11.4	14.3	5.1	.820
Loudness	3.2	6.3	7.7	5.1	.720

Comparisons between groups performed with the Kruskal-Wallis test or χ^2 test as applicable.
CAPE-V = Consensus Auditory-Perceptual Evaluation of Voice; CM = conservative management only; CM + PM = conservative management followed by permanent medialization; IL = temporary injection laryngoplasty only; IL + PM = temporary injection laryngoplasty followed by permanent medialization.

TABLE V.
Normalized Glottic Gap Areas and Average CAPE-V Measurements at Presentation and Last Follow-Up IL Versus Any PM Groups.

	Presentation			Last Follow-Up		
Variable	IL (n = 19)	Any PM (n = 28)	*P* Value	IL (n = 19)	Any PM (n = 28)	*P* Value
Normalized glottic gap area	4.66	5.27	.39	1.21	1.14	.86
CAPE-V characteristic						
Overall severity	53.4	59.6	.43	25.4	31.2	.11
Roughness	23.9	21.4	.89	13.3	18.3	.03*
Breathiness	42.7	51.0	.34	12.5	13.0	.14
Strain	20.4	17.1	.95	14.0	12.9	.99
Pitch	27.3	25.6	.61	14.3	10.2	1.00
Loudness	29.2	36.6	.36	7.7	6.1	.67

Comparisons between groups performed using a χ^2 test.
*Effect with $P < .05$.
CAPE-V = Consensus Auditory-Perceptual Evaluation of Voice; IL = temporary injection laryngoplasty only; PM = permanent medialization.

presentation (3.6 months, $P = .018$). Furthermore, surgery was noted to be the etiology of vocal fold paralysis in all patients in the IL + PM group, which reached statistical significance when compared to the other treatment groups (8/8 surgery, $P = .003$). These differences among groups may suggest that younger patients are more amenable to seeking treatment quicker and opting for more invasive treatment strategies when compared to an older cohort of patients. Given the iatrogenic nature of these injuries, it is also possible that there remained a tendency on the part of our surgical team to intervene with voice rehabilitation surgery earlier in the postoperative course than would have otherwise been performed for injuries not sustained due to surgery. As noted in Table I, time to presentation was found to be greater for those patients opting for conservative treatment alone when compared to those patients who underwent an IL procedure. This discrepancy likely reflects the practice tendency to offer IL early following the diagnosis of UVFP in an effort to ameliorate symptoms while waiting for possible return of function. In addition, patients who opted for conservative treatment may have been more likely to do so given the length of time these patients had been living with, and potentially compensating for, their symptoms of UVFP.

Despite these differences in descriptive characteristics, however, we found that patients are similar across groups with respect to NGGA and perceptual voice scores as judged by CAPE-V analyses at presentation. As shown in Table III, the NGGA for each group is similar at presentation. These similarities among groups persist with respect to average CAPE-V scores at presentation; however, the CM group was noted to have slightly improved average scores for Breathiness (10.7, $P = .007$) and Loudness (2.8, $P = .018$). Thus, although the glottic gap does not seem different for the CM patients when compared to the other subjects, their perceptual voice quality may be better than the other three groups. This finding perhaps explains why these patients opted against an IL or PM procedure in lieu of more conservative management of their injury.

Interestingly, whereas patients were noted to have overall improvements in NGGA and average CAPE-V scores at last follow up in those groups who underwent intervention for their UVFP, we found no significant differences amongst groups at last follow-up with respect to NGGA or average CAPE-V scores (Table IV). This would suggest that, regardless of treatment strategy, all of our patients appear similar with regard to glottic competence, as judged by glottic gap and perceptual voice at long-term follow-up. A similar result is observed when we compare the IL group to those patients who went on to receive a PM procedure (IL + PM and CM + PM groups, collectively referred to as Any PM). Again, those patients in the IL group appear similar to those in the Any PM group at last follow-up. The lone exception to this finding is that the IL group appears to have slightly improved average CAPE-V scores for Roughness (13.3 vs. 18.3, $P = .03$) when compared to the Any PM group at last follow-up (Table V).

These results suggest that early IL with an agent designed to have a temporary duration of action may result in sustained improvement in glottic competence and perceptual voice in a certain cohort of patients. Although our group and others have previously demonstrated an association between IL and decreased rates of subsequent PM, a recent retrospective study by Francis et al. found that IL had no effect on the need for definitive framework surgery in a cohort of 633 patients.[5] This study is important in that it cautions readers to consider selection bias when interpreting the association between IL and need for subsequent PM procedures, given that many patients with UVFP may achieve spontaneous recovery of vocal fold function despite intervention. To mitigate against this potential bias in the present study, we excluded any patients with spontaneous return of vocal fold function noted during treatment. Furthermore, patients included in the present study had a median time to last follow-up of 22.3 months (range, 19.4–28 months), which is well-beyond the generally accepted time frame for spontaneous return of function following injury. In this light, we feel that the present study still bears relevance with respect to how to counsel patients with UVFP with unknown potential for functional recovery.

In this study, we used the measurement of NGGA as an objective measure of glottic closure during phonatory tasks. First reported by Omori and colleagues, this measurement reports the maximum glottic gap area appreciated during sustained phonation when referenced against the length of the membranous true vocal fold.[12] Although there remains slight subjectivity while capturing the moment of maximum glottic closure during videolaryngostroboscopy, this method has gained popularity as a relatively robust and objective tool to assess maximum glottic closure.[13,14] Fang et al. recently utilized NGGA as a method by which to predict which patients might go on to require PM for a diagnosis of UVFP. Their findings suggest that an NGGA cutoff value of 7.36 resulted in 85.7% sensitivity and 80% specificity for predicting future PM.[13] Of note, our average NGGA at initial presentation was noted to be well below this particular cutoff. Given our results of similar NGGA across groups at presentation and last follow-up, our study would suggest that treatment recommendations may not be based on the size of the glottic gap alone.

In addition to objective measures of glottic competence, we also considered perceptual evaluation of voice by way of CAPE-V analyses. To evaluate the reliability of this semiobjective metric, intra- and inter-rater reliability was calculated based on the Pearson correlation method. As demonstrated in Table II, our intrarater reliability for each of our two independent SLPs was judged to be good based on Pearson correlation coefficients $r > 0.5$ across all characteristics with the exception of Strain for SLP 2. This is consistent with other studies demonstrating Strain as the least reliable characteristic studied in this particular grading battery.[9,10] In addition, inter-rater reliability was noted to be fairly strong with 4/6 CAPE-V characteristics having correlation coefficients $r > 0.5$. The most disparate ratings were found when comparing scores for Roughness and Strain, which may be more difficult to rate in this patient cohort when compared to other voice characteristics such as Breathiness or Loudness.

Although the CAPE-V tool is a commonly employed and validated metric for measuring perceptual voice, we would caution readers with respect to interpretation of Loudness scores in the present study. This voice characteristic is contingent upon a number of factors that might influence observer perception. As addressed in our Materials and Methods section above, some of these factors include the acoustics of the room in which voice recordings are performed, ambient noise that may be present during time of recording, and nonstandardized distance between the mouth and microphone during phonatory tasks. In this study, it should be noted that we could not reliably control for these specific factors. Although Loudness was only noted to be a perceived difference in our CM group relative to other patients upon presentation, these results should, nonetheless, be interpreted in this context as a limitation of our use of the CAPE-V grading system.

Other limitations of the present study include the retrospective nature and relatively small sample size. As with any retrospective review, one must consider the possibility of selection bias and patients who are lost to follow-up. In the present study, only approximately 20% of patients identified with a diagnosis of UVFP met all inclusion criteria for review. Notably, the majority of patients who were excluded from analysis was due to insufficient follow-up. We accept this selection bias given that the primary objective of this study was to look at long-term comparisons amongst patients, and thus favor more stringent inclusion criteria. The number of patients excluded, however, translates to a failure to capture both patients who may have had less than satisfactory results following an intervention, as well as those who may have gone on to attain spontaneous return of function or achieve adequate glottic compensation without further treatment. Although this selection bias exists, we still feel that the findings herein are valuable to consider when counseling patients with UVFP.

In addition, as patients in this retrospective study were not randomly assigned to treatments, decisions to offer a particular treatment may be biased and could possibly influence our results. This study also highlights the need for more robust and universally accepted metrics for evaluating glottic competency and voice outcomes in patients with UVFP. Although use of the NGGA remains a commonly employed method for quantitatively assessing glottic closure, its utility in informing functional outcomes and prognosis remains less clear. In addition, although often used as a validated instrument for grading perceptual evaluation of voice, the CAPE-V tool may be subject to interpretation bias and confounders with respect to how voice samples are acquired. Despite these shortcomings, we feel that the acceptable intra- and inter-rater reliability data substantiate the use of these metrics in the present study. Although our study supports the notion that IL may result in long-term benefit similar to patients who receive PM, further studies are needed to better delineate the mechanisms driving this result and patient selection criteria.

CONCLUSION

Our results support some previous findings that IL with an agent designed to have a temporary duration of action may have more lasting and favorable outcomes than initially postulated. Furthermore, we demonstrate that treatment decisions may not be predicated on the size of the glottic gap alone given the homogeneity in NGGA among groups at presentation. Lastly, we demonstrate that there remains a subset of patients in whom temporary IL yields results that are equivalent to those obtained by PM at long-term follow-up. More studies are needed to help elucidate the mechanisms driving this effect and to assist with selection of the optimal candidate for intervention.

Acknowledgments

The authors acknowledge Chengshi Jin, PhD, MS, and Iryna Lobach, PhD, for their contributions to the statistical analyses performed in this study.

BIBLIOGRAPHY

1. Courey MS. Homologous collagen substances for vocal fold augmentation. *Laryngoscope* 2001;111:747–758.
2. Friedman AD, Burns JA, Heaton JT, Zeitels SM. Early versus late injection medialization for unilateral vocal cord paralysis. *Laryngoscope* 2010;120:2042–2046.
3. Yung KC, Likhterov I, Courey MS. Effect of temporary vocal fold injection medialization on the rate of permanent medialization laryngoplasty in unilateral vocal fold paralysis patients. *Laryngoscope* 2011;121:2191–2194.
4. Prendes BL, Yung KC, Likhterov I, Schneider SL, Al-Jurf SA, Courey MS. Long-term effects of injection laryngoplasty with a temporary agent on voice quality and vocal fold position. *Laryngoscope* 2012;122:2227–2233.
5. Francis DO, Williamson K, Hovis K, et al. Effect of injection augmentation on need for framework surgery in unilateral vocal fold paralysis. *Laryngoscope* 2016;126:128–134.
6. Carroll TL, Rosen CA. Long term results of calcium hydroxylapatite for vocal fold augmentation. *Laryngoscope* 2011;121:313–319.
7. Young VN, Gartner-Schmidt J, Rosen CA. Comparison of voice outcomes after trial long-term vocal fold augmentation in vocal fold atrophy. *Laryngoscope* 2015;125:934–940.
8. Carroll TL, Rosen CA. Trial vocal fold injection. *J Voice* 2010;24:494–498.
9. Kempster GB, Gerratt BR, Verdolini Abbott K, Barkmeier-Kraemer J, Hillman RE. Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. *Am J Speech Lang Pathol* 2009; 18:124–132.
10. Zraick RI, Kempster GB, Connor NP, et al. Establishing validity of the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V). *Am J Speech Lang Pathol* 2011;20:14–22.
11. Rasband WS. ImageJ. U. S. National Institutes of Health, Bethesda, Maryland, USA. Available at: https://imagej.nih.gov/ij. Accessed 2015–2016.
12. Omori K, Kacker A, Slavit DH, Blaugrund SM. Quantitative videostroboscopic measurement of glottal gap and vocal function: an analysis of thyroplasty type I. *Ann Otol Rhinol Laryngol* 1996;105:280–285.
13. Fang TJ, Pei YC, Li HY, Wong AM, Chiang HC. Glottal gap as an early predictor for permanent laryngoplasty in unilateral vocal fold paralysis. *Laryngoscope* 2014;124:2125–2130.
14. Wang CC, Chang MH, Jiang RS, et al. Laryngeal electromyography-guided hyaluronic acid vocal fold injection for unilateral vocal fold paralysis: a prospective long-term follow-up outcome report. *JAMA Otolaryngol Head Neck Surg* 2015;141:264–271.