

2025

Measures of supraglottic constriction: comparing visual-perceptual and quantitative measures in trans and gender-nonconforming individuals taking testosterone

<https://hdl.handle.net/2144/50725>

"Downloaded from OpenBU. Boston University's institutional repository."

BOSTON UNIVERSITY
SARGENT COLLEGE OF HEALTH AND REHABILITATION SCIENCES

Thesis

**MEASURES OF SUPRAGLOTTIC CONSTRICTION:
COMPARING VISUAL-PERCEPTUAL AND QUANTITATIVE MEASURES
IN TRANS AND GENDER-NONCONFORMING INDIVIDUALS
TAKING TESTOSTERONE**

by

JORDAN E. KIMMEL

B.S., Saint Michael's College, 2022

Submitted in partial fulfillment of the
requirements for the degree of
Master of Science

2025

© 2025 by
JORDAN E. KIMMEL
All rights reserved

Approved by

First Reader

Cara E. Stepp, Ph.D.
Professor of Speech, Language & Hearing Sciences
Sargent College of Health and Rehabilitation Sciences

Professor of Biomedical Engineering
College of Engineering

Professor of Otolaryngology – Head and Neck Surgery
Aram V. Chobanian & Edward Avedisian School of Medicine

Second Reader

Daniel P. Buckley, M.S., CCC-SLP
Lecturer in Speech, Language & Hearing Sciences
Sargent College of Health and Rehabilitation Sciences

Assistant Professor of Otolaryngology – Head and Neck Surgery
Aram V. Chobanian & Edward Avedisian School of Medicine

Third Reader

Rachael E. Kammer, M.S., CCC-SLP, BCS-S
Clinical Assistant Professor of Speech, Language & Hearing Sciences

Fourth Reader

Allison Aaron, M.S., CCC-SLP
Doctoral Candidate in Speech, Language & Hearing Sciences
Stepp Lab for Sensorimotor Rehabilitation Engineering

ACKNOWLEDGMENTS

This thesis would not have been possible without the incredible support I have received throughout my graduate career. Firstly, thank you to Dr. Cara Stepp for welcoming me into her lab from the moment she learned of my admission. You are the reason I came to BU, and the opportunity to not only work in your lab but be an integral part of the life-changing research being done on transgender and gender-nonconforming (TGNC) populations has been one of the most exciting parts of this experience. Your unwavering support of my research-related learning and voice-related endeavors has been invaluable. I greatly appreciate the wealth of knowledge that you have provided as well as your deep commitment to researching a population that is historically under-represented while also including TGNC individuals in the research process.

I also would not have been able to complete this paper without the immense help of Allison Aaron who pushed me to be my best from the very start. Your passion for and knowledge of research has allowed me to better understand concepts and skills that were not only necessary for this project but will shape my career as an SLP. You have shown that you genuinely care for this population and are always willing to learn how to best support our outcomes.

I would also like to thank Daniel Buckley and Rachael Kammer for serving as members of my committee. Daniel, your passion for voice is palpable and has allowed me to learn more about my own passion for voice. Your support as a professor, clinical supervisor, and thesis committee member has shaped my knowledge and skills for voice care immensely, which has been crucial to my understanding of how I can support my

own community through my practice. I cannot thank you enough for aiding in the opportunity to pursue what brought me to this program in the first place. Rachael, your support of this project and willingness to learn and understand more about this population is so appreciated. Your knowledge and expertise have gone far beyond being a professor and clinical supervisor. It is clear that you care tremendously about student success as a whole and I have enjoyed getting to know you and work with you.

The number of individuals in the lab that have either contributed to this project or contributed to my success is innumerable. First, thank you to Jose Rojas for creating the MATLAB scripts and always being willing to troubleshoot and make updates. Thank you also to Taylor, Jade, and Misora for both completing data analysis and acting as reliability, and to Allison, Daniel, Kim, Sam, and Nichole for the *many* VALI ratings you all completed. Last, but not least, thank you to the friends I've made in the lab. I've appreciated all our time spent together, whether it was in the lab or at the beach/hockey rink.

Finally, I would be remiss if I did not thank my friends and family for all their support throughout this past year and a half. To my friends also completing a thesis this spring, we did it! This process is not easy, and I am thankful for the camaraderie and support we have all had for each other and the ability to learn from each other. To the friends that I have made in this program, we also did it! I cannot thank you all enough for always encouraging and supporting me in every way. To my mom, dad, and brother, I would not be where I am today without having your lifelong support. Thank you for always giving me a place to come home to that was nothing but loving and encouraging.

**MEASURES OF SUPRAGLOTTIC CONSTRICTION:
COMPARING VISUAL-PERCEPTUAL AND QUANTITATIVE MEASURES
IN TRANS AND GENDER-NONCONFORMING INDIVIDUALS
TAKING TESTOSTERONE**

JORDAN E. KIMMEL

ABSTRACT

Purpose: The purpose of this study was to determine which quantitative measures of supraglottic constriction were most closely related to visual-perceptual measures of supraglottic activity in transgender and gender-nonconforming (TGNC) people who are using testosterone (T) as a form of gender-affirming hormone therapy (GAHT). Specifically, we wanted to explore whether instances of vocal hyperfunction as measured by supraglottic constriction, occur alongside acoustic changes in fundamental frequency (f_0), which may be associated with structural laryngeal changes.

Method: Eleven participants who were assigned female at birth (AFAB) and taking T for the first time, were required to complete 13 monthly visits to the lab (baseline prior to starting T and then monthly for the year following onset) and be willing to undergo flexible laryngoscopy. Acoustic recordings and laryngoscopic images were analyzed to determine at which visit the largest monthly decrease in f_0 occurred in semitones (ST) and to evaluate anteroposterior (AP) and mediolateral (ML) constriction at baseline, 1-month following the start of T, and at the visit following the largest monthly decrease in f_0 using quantitative and visual-perceptual (Voice-Vibratory Assessment With Laryngeal Imaging; VALI) measures of supraglottic constriction. Linear mixed-effects models were

utilized to determine if the independent variables (quantitative measures) predicted the VALI. β values were used to determine significant effects.

Results: This study found that in the AP direction, two out of two quantitative measures significantly predicted VALI outcomes, whereas in the ML direction, only two out of four quantitative measures significantly predicted VALI outcomes. There was more variation across the range of VALI ratings in the AP direction as compared to the little variation found in the ML direction.

Conclusion: These results suggest that AP measures can be used to determine variation in constriction clinically, but that ML measures should be used with caution. The results also help to explain more about what is occurring in the larynx during large and fast changes to its structure in individuals who were AFAB and take GAHT with T. Rather than variation occurring at an expected timepoint (following a large decrease in f_0), variation occurred throughout the first year on T. These findings, in conjunction with more research regarding constriction and self-reported symptoms, can aid in individualizing evaluation and treatment of voice in TGNC people on T.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	iv
ABSTRACT	vi
TABLE OF CONTENTS	viii
LIST OF TABLES	x
LIST OF FIGURES	xi
Introduction	1
Acoustic changes with testosterone as gender-affirming hormone therapy	1
Voice symptoms on T.....	2
Measures associated with hyperfunctional voicing.....	3
Voice-Vibratory Assessment With Laryngeal Imaging	6
Quantitative measures of supraglottic constriction	8
Methods	13
Data Collection.....	13
Data Analysis.....	16
Statistical Analysis	20
Results	21
Outcome measures as a function of time.....	21
Reliability	23
Comparing the VALI to quantitative measures.....	24
Discussion.....	27
AP direction.....	29

ML direction.....	31
Clinical implications.....	33
Limitations and future directions.....	36
Conclusion.....	37
BIBLIOGRAPHY	38
CURRICULUM VITAE	42

LIST OF TABLES

Table 1. Participant Information.....	14
---------------------------------------	----

LIST OF FIGURES

Figure 1. Rainbow Passage fundamental frequency (f_0).	17
Figure 2. Visual-perceptual (VALI) measures of supraglottic constriction.	22
Figure 3. Two quantitative measures of AP supraglottic constriction.	22
Figure 4. Four quantitative measures of ML supraglottic constriction.	23
Figure 5. Quantitative measures that predict VALI measure outcomes.....	26

Introduction

Acoustic changes with testosterone as gender-affirming hormone therapy

The most common way for transgender and gender-nonconforming (TGNC) assigned female at birth (AFAB) individuals to achieve voice masculinizing effects is through gender-affirming hormone therapy (GAHT) with testosterone (T). Over the course of one year on T, many of these individuals experience large acoustic changes in fundamental frequency (f_0), the acoustic correlate of pitch. On T, f_0 typically becomes lower, and voice masculinizing effects are associated with a lower f_0 (Hancock et al., 2017). This change has been documented by several studies to occur over a relatively short period of time (Cler et al., 2020; Damrose, 2009; Deuster et al., 2016; Irwig et al., 2017; van Borsel et al., 2000). One study reported that acoustic changes most often occur within or around 6 months of treatment with T (Irwig et al., 2017). The steepest changes have been documented between 2 and 4 months into treatment (Cler et al., 2020; Damrose, 2009; Deuster et al., 2016; van Borsel et al., 2000).

It is unclear what factors might be associated with these acoustic changes, which is why additional research is needed to better understand both what might be causing acoustic changes and whether compensatory voicing behaviors (i.e., physiological or functional changes) might occur due to the presumed structural/anatomical changes in this population. Given that the f_0 of an individual's voice is related to structural factors of the larynx, such as their vocal fold length and thickness (Ferrand, 2011), it is assumed that these structures change drastically with exposure to T (Azul et al., 2017). However, there is a gap in knowledge, as laryngeal structural changes resulting from taking T are

scarcely documented and mostly speculated. There is also little known about how individuals might vocally adjust and adapt to these supposed structural changes.

Voice symptoms on T

There is evidence to suggest that TGNC individuals have voice complaints and symptoms. Commonly reported symptoms include vocal fatigue, negative change in voice quality, and reduced control of vocal gender presentation (Azul et al., 2018; Cler et al., 2020; Nygren et al., 2016). Azul and colleagues (2018) noted that in a group of 13 TGNC AFAB individuals who were currently taking T, seven participants noted control of voice presentation as “difficult” or “at times difficult,” and one participant noted “smoky voice quality” as a problem. Nygren and colleagues (2016) found that 12 out of 50 transgender men who had started T sought out voice therapy after the start of treatment with T. These individuals pursued therapy due to voice complaints such as vocal fatigue, vocal instability, strained voice quality, hoarseness, and difficulty projecting the voice. In a longitudinal study by Cler and colleagues (2020) on one participant taking T, data were collected three times in two weeks prior to starting T, and then every two weeks after starting T over the course of one year. The participant expressed fatigue while using his voice and difficulty singing in a high pitched, whispered voice as 1–3 on a scale of 1–10 at baseline, increasing to 4–8 during the middle period of the year, then decreasing back to 3–5 towards the end of 1 year. This pattern of increased fatigue is aligned with the participant’s largest drop in f_0 (between months 3–4 and months 5–6). This finding suggests that the decrease in f_0 may be leading to further functional changes, thus precipitating voice symptoms. Despite these findings,

we do not know what might be causing these self-reported voice symptoms. They may be a result of physiological voice changes or changes to the function of the voice in which the muscles of the larynx are being overworked (hyperfunction; Azul et al., 2018; Cler et al., 2020; Nygren et al., 2016; Oates & Winkworth, 2008). The combination of self-reported voice symptoms and suspected physiological changes motivate the investigation of hyperfunctional voicing following large decreases in f_0 for TGNC individuals taking T.

Measures associated with hyperfunctional voicing

Given the evidence for reported voice symptoms on T, it would be worthwhile to investigate quantitative measures associated with vocal hyperfunction in TGNC individuals taking T. Vocal hyperfunction is defined as “excessive peri-laryngeal musculoskeletal activity during phonation” (Hillman et al., 2020). Large and fast structural changes in the larynx may induce vocal hyperfunction in TGNC people taking T, as their body is learning to adapt to these changes in new and different ways. Vocal fatigue, which may also lead to vocal hyperfunction, is often considered to be a result of prolonged voice use. Strategies that are used to compensate for and adapt to changes in/use of the voice are also thought to lead to hyperfunctional voicing (Solomon, 2008). Thus, adaptation of the laryngeal musculature in periods of considerable structure change may precede vocal hyperfunction.

There are several ways to assess vocal hyperfunction. These include auditory-perceptual evaluation, acoustic measures of vocal function, and laryngeal endoscopy (Patel et al., 2018). Auditory-perceptual evaluation can be used to measure aberrant vocal quality. Certain acoustic measures are associated with dysphonia, which can be described

as an impairment in the production of the voice that can correspond to both absent vocal fold pathology and present vocal fold pathology (either structural [e.g., nodules] or neurogenic [e.g., paralysis] in nature; Sachdeva & Shrivastava, 2018). Measures of acoustics include cepstral peak prominence, low-high spectral ratio, and harmonics-to-noise ratio. These acoustic measures can be used alongside measures of laryngeal endoscopy to understand more about potential hyperfunctional voicing patterns.

Video-stroboscopic exams are typically completed using a trans-nasal fiberoptic laryngeal endoscope to visualize the physiology of the larynx (vocal folds and tissue above them) during phonation and to measure supraglottic constriction (adduction of the false vocal folds and compression of the arytenoid cartilages to the petiole of the epiglottis; Stager et al., 2001). Supraglottic constriction can be measured using laryngeal endoscopy, visualizing the constriction of the tissue above the vocal folds (Patel et al., 2018). The constriction of the structures above the glottis is thought to represent compensation within the larynx, which can lead to voice symptoms and difficulty communicating (Fernandez et al., 2020; Stager et al., 2000). Laryngoscopy has been used to assess vibratory findings in one AFAB individual who was undergoing GAHT with T and there were no atypical findings (besides an intubation granuloma from an unrelated surgery that was visible from weeks 34–44 of treatment), though constriction was not assessed in this study (Cler et al., 2020).

Supraglottic constriction has been studied in transgender people who were assigned male at birth (AMAB; Palmer et al., 2012). Palmer and colleagues (2012) sought out to determine whether there were signs of vocal misuse or hyperfunction in the

vocal fold activity of AMAB individuals who were attempting to produce a more feminine voice. They used modified visual-perceptual measures of supraglottic activity, finding varying degrees of supraglottic constriction in all participants as well as that 67% of this group reported voice complaints (Palmer et al., 2012). Supraglottic constriction has not yet been investigated in TGNC people who were AFAB. This is important to note, as we are predicting large structural changes in AFAB people, whereas AMAB people are manipulating the structures they have. This is because estrogen does not feminize the voice (Cler et al., 2020; Palmer et al., 2012). Evidence that manipulation of structures may lead to variation in supraglottic constriction and voice complaints in AMAB TGNC individuals may support our interpretation of how large and fast structural changes affect how AFAB TGNC individuals use their voices (Palmer et al., 2012).

Despite laryngologists' use of supraglottic constriction to make clinical diagnoses, there is conflicting evidence about the association between supraglottic constriction and vocal hyperfunction in the literature. This conflict may be explained by the variation in the ways that constriction is both measured and interpreted. Typically, constriction is measured in two directions, anteroposteriorly and mediolaterally. Anteroposterior (AP) constriction is defined as how much of the vocal folds can be seen during speech tasks (Fernandez et al., 2020). Mediolateral (ML) constriction is defined as whether the false vocal folds can be seen and if so, whether they are in contact with each other during speech tasks (Fernandez et al., 2020). Studies by Shembel and colleagues (2023) and Dabirmoghaddam and colleagues (2021) determined that ML constriction can be present in both control and disorder groups, aiding in the explanation that ML constriction might

not be indicative of muscle tension dysphonia or vocal hyperfunction (Dabirmoghaddam et al., 2021; Shembel et al., 2023). There was however, significantly more severe constriction in the disorder group of Shembel and colleagues (2023). On the other hand, AP constriction may be more indicative of hyperfunction (Dabirmoghaddam et al., 2021). Some researchers posit that supraglottic constriction is not a perfect measure for understanding the differences in laryngeal activity in those with normal voicing compared to hyperfunctional voicing (Dabirmoghaddam et al., 2021). Although there is no standardized way to measure supraglottic constriction in either direction and conflicting evidence on its association with vocal hyperfunction; we do not yet know whether there are quantitative measures which can determine a change in constriction after predicted large structural changes that may correlate with hyperfunctional voicing.

Voice-Vibratory Assessment With Laryngeal Imaging

Currently, the most commonly used visual-perceptual tool to measure supraglottic constriction is the Voice-Vibratory Assessment With Laryngeal Imaging (VALI; Poburka et al., 2017). The VALI form is a visual-perceptual rating form that involves judgements from a single rater. The VALI includes 11 parameters; one parameter is termed supraglottic activity and is defined as “constriction of the supraglottic structures” (Poburka et al., 2017). Raters are instructed to evaluate these parameters according to their definitions. Videostroboscopic recordings are captured of the participants’ sustained /i/ at typical pitch and loudness. Raters measure supraglottic activity by determining the qualitative amount of constriction through visuo-perceptual ratings on both the AP and ML planes on a scale of 0–5. In a study that evaluated the reliability of the VALI, thirty

participants with dysphonia provided video samples for the study and raters completed the VALI form for these samples. In the study, the results of the VALI ratings showed that ML supraglottic activity ratings should be interpreted with caution due to the reduced sample size of rated samples compared to other parameters. Although this was the case, inter-judge reliability for ML supraglottic activity showed very strong intraclass correlation coefficients, suggesting that there was consistency across raters. However, intra-judge reliability was not significant for either parameter of supraglottic activity (AP or ML) for stroboscopy (Poburka et al., 2017). These findings suggest that the VALI supraglottic activity parameter may not be as reliable, compared to the VALI ratings for other parameters. Further, it involves judgements from a single rater when used in clinical practice, so it may not adequately measure sensitivity to change across multiple time-points or raters.

To objectively measure supraglottic constriction across multiple conditions (time, status change, treatment), it is crucial to evaluate quantitative measures for high sensitivity to change in differences of supraglottic constriction. These conditions (time, status change, treatment) may include measuring constriction before or after events such as voice therapy, surgery, using a therapy tool, or using T as GAHT. A few quantitative measurements of supraglottic constriction in patients with voice disorders have been designed, measuring either AP or ML constriction, or both, using several different analysis processes (Behrman et al., 2003; Han et al., 2023; Ogawa et al., 2013; Stager et al., 2001). To date, there is no consensus as to best practices for quantifying supraglottic constriction. Evaluating these varied measures and comparing them with the VALI form

is necessary to determine how to best measure change in supraglottic constriction for those with voice symptoms. Specifically, it can help us to evaluate hyperfunctional changes in TGNC individuals taking T and to best serve TGNC individuals on T who seek out voice therapy.

Quantitative measures of supraglottic constriction

Behrman and colleagues (2003) examined supraglottic constriction in 40 patients with dysphonia and 40 typical controls. This study compared constriction measures between groups. They completed videostroboscopic recordings of sustained /i/ phonation. Using images from these recordings, they measured the length and width of the visible vocal folds and the area of the laryngeal outlet, defined as “the area above the glottis bounded by the petiole of the epiglottis anteriorly, arytenoids posteriorly, and the ventricular folds laterally” (Behrman et al., 2003). They normalized the length and width of the vocal folds to the laryngeal outlet to determine the level of AP or ML supraglottic constriction. AP and ML constriction were shown to be reliably measured, due to the differences between images for both normal controls and dysphonic patients not being statistically significant. It was found that relative AP constriction was significantly greater in the dysphonic group, with overlap between groups in the lower values of constriction. There were no significant differences between groups with regard to relative ML constriction (Behrman et al., 2003). These findings suggest that there are constriction differences between typical controls and dysphonic patients, but only in the AP direction.

Han and colleagues (2023) examined supraglottic constriction in 33 subjects with primary muscle tension dysphonia and 30 typical speakers. This study compared

constriction measures between the two groups and across voicing and non-voicing tasks, exclusively on the ML plane. Data were recorded using a flexible laryngoscope during phonation of /i/ and maximum inspiration during quiet breathing both before and after various voicing tasks. Images were taken from these videos, on which five points were plotted, equally spaced apart from each other on each false vocal fold then plotting these along the y axis of a coordinate grid with the anterior most point at (0,0) and the posterior most point also on the y axis. This figure was fit to a curve to determine whether it was positive (i.e., hyperfunctional) or negative (i.e., relaxed). There was a significant difference in false vocal fold curvature between the voice disorder and control groups during maximal inspiration, but no significant difference between groups during normal voicing with /i/, which led authors to interpret that hyperfunctional contour of the false vocal folds during quiet breathing is more indicative of a hyperfunctional voice disorder than a hyperfunctional contour during phonation. This measure is potentially more sensitive to change in quiet breathing than it is in phonation (Han et al., 2023).

Ogawa and colleagues (2013) examined supraglottic constriction in 23 patients with muscle tension dysphonia and 15 age-matched individuals with sleep apnea but no laryngeal abnormalities. This study compared constriction measures between the two groups and across voicing and non-voicing tasks. Data were recorded using a flexible laryngoscope during sustained phonation of /i/ and the maximal opening of the vocal folds during inspiration. The authors took images from these videos and measured the AP distance of the visible vocal folds and medial width from the midpoint of one false vocal fold to the midpoint of the other false vocal fold. They normalized these measurements to

the mean length of the vocal folds at maximal inspiration, measured from the anterior commissure to the bilateral vocal processes. Participants were also asked to hum /m/ several times. This measurement of supraglottic constriction showed that humming reduced both AP and ML constriction in both dysphonic and nondysphonic patients. Findings show that this constriction measure was sensitive to change after humming in both AP and ML measurements (Ogawa et al., 2013).

Stager and colleagues (2001) examined supraglottic constriction in 12 participants with complaints of vocal fatigue or diagnosis of vocal fold nodules, compared to 12 controls. This study compared constriction measures between the two groups and across voicing and non-voicing tasks. Using a flexible laryngoscope, they recorded the points of maximum and minimum constriction during both adduction and abduction of the vocal folds for many different tasks, including quiet breathing and phonation. Using still images taken from the video data, the AP distance was measured as “the length in pixels from the midpoint of the posterior commissure to the junction of the two vocal folds at the anterior commissure (if visible) or the petiole of the epiglottis if it was covering the anterior commissure” (Stager et al., 2001). The vocal fold length was measured as “the length in pixels of the visible portion of the medial edge of each [vocal fold] from the anterior commissure until the medial edge intersects with the arytenoid cartilage” (Stager et al., 2001). The vocal fold area was measured as “the area in pixels of the visible portion of each [vocal fold]” (Stager et al., 2001). AP constriction was calculated by using the AP distance, and ML constriction was calculated as the vocal fold width (by dividing the vocal fold area by the vocal fold length). These measurements were

normalized by dividing the AP distance for a specific task by the greatest AP distance from any task and dividing the vocal fold width for a specific task by the greatest vocal fold width from any task. There were no significant differences between groups found for ML constriction, but there were significant differences between groups found for AP constriction (Stager et al., 2001).

All of these studies compare constriction between patient populations and a control group. Three of the four studies also compared constriction across various tasks within a single session. The studies all used different approaches to quantify constriction (i.e., different sample sizes, different vocal/non-vocal tasks, different methods for measuring constriction), with conflicting findings. Behrman et al. (2003) and Stager et al. (2001) reported that AP constriction may be more quantifiable than ML constriction. Han et al. (2023) suggested that hyperfunctional voicing may be more prominent during quiet breathing as compared to phonation. Ogawa et al. (2013) stated that humming may best reduce hyperfunctional effects. However, none of these studies investigated whether the measures were sensitive to change before and after structural changes due to T. A tool that quantitatively measures vocal changes in TGNC individuals who take T as GAHT would be useful in understanding more about the time-course of structural and acoustic changes to expect for these individuals throughout the first year on T. To better understand how functional changes are associated with known acoustic changes and assumed structural changes in TGNC individuals taking T, quantitative and visual-perceptual measures of supraglottic constriction should be compared with one another

and across multiple time-points throughout the first year of taking T (which is when the largest acoustic changes are expected; Cler et al., 2020).

The purpose of this study was to determine which quantitative measures of supraglottic constriction are most closely related to visual-perceptual measures of supraglottic activity in TGNC people who are using T as a form of GAHT. Specifically, we wanted to explore whether instances of vocal hyperfunction as measured by supraglottic constriction, occur alongside acoustic changes in f_0 , which may be associated with structural laryngeal changes. Large and fast changes in f_0 may imply large structural changes, which may lead to supraglottic constriction. This is due to adaptation needs in the supraglottic area, which could induce constriction, eventually leading to voice problems. During the first year of treatment with T as GAHT, there are large amounts of variability. We used this variability within speakers to observe if previously determined quantitative measures compare to visual-perceptual measures. Video-stroboscopic exams were completed using a trans-nasal fiberoptic laryngeal endoscope to visualize the physiology of the larynx during phonation and to measure supraglottic constriction. Baseline stroboscopic exams were collected before participants start treatment and were compared to stroboscopic exams collected one month after starting treatment (as a control) and during the visit following the participants' largest decrease in f_0 . Stroboscopic data were used to observe changes in supraglottic constriction which could be an indicator of vocal hyperfunction. The VALI was compared to objective measures of supraglottic constriction, which use quantitative protocols (Behrman et al., 2003; Han et al., 2023; Ogawa et al., 2013; Stager et al., 2001). We expected that existing quantitative

metrics of AP and ML constriction would be related to the VALI metrics of AP and ML supraglottic constriction. We hypothesized that certain quantitative measures would better predict VALI outcomes as compared to others. Specifically, the measure designed by Behrman and colleagues would best predict VALI outcomes, because it measures the AP and ML constriction levels compared to the laryngeal outlet in a way that is similar to VALI ratings.

Methods

Data Collection

Participants

Participants included TGNC individuals who were AFAB and were about to begin GAHT with T. These individuals did not have a previous history of taking T. Enrolled participants were at least 18 years old. We collected information on administration method of T, dosage taken, and schedule of dosage (Table 1). Administration methods included injection (subcutaneous or intramuscular) and transdermal (gel, patch, etc). To be included in this study, participants must have continually been taking T, participated in a baseline visit prior to starting T and 12 monthly visits (at 1-month intervals after starting T) to the lab at Boston University for one year, and been willing to undergo flexible laryngoscopy. Amount of time on T was measured in months (month 1 = 1 month post-T, month 2 = 2 months post-T, etc.). Informed consent was obtained as per the Boston University Institutional Review Board.

During the baseline visit, participants must have passed a hearing screening to ensure they were at the capacity to provide informed consent. This screening was

completed with over-ear audiometer headphones and pulsed tones at 250, 500, 1000, 2000, 4000, and 8000 Hz (ASHA, 2005). Each month, surveys were completed to document participants' self-reported voice and gender identity. These surveys included the Voice Related Quality of Life (VRQOL) and the Transgender Self Evaluation Questionnaire (TSEQ). At the initial, baseline visit, participants were asked if they had ever participated in voice training or therapy (whether gender-related or not). Only 2 out of 11 (P2 and P3) reported yes.

Table 1. Participant Information. Dosage and method of taking T are noted to further describe variation between participants, but this variation was not factored in to analysis of constriction changes.

PT ID#	Age (at month 0)	Gender Identity (at month 0)	Method of Taking T	T dosage (range throughout 1 year)	Largest f_0 decrease in semitones (ST)
P1	21	Nonbinary	Injection	50 mg/week–140 mg/2 weeks	Between month 1 and 2
P2	18	Genderqueer	Gel	12.5–50 mg/day	Between month 3 and 4
P3	18	Man	Injection	40–50 mg/week	Between month 1 and 2
P4	25	Transmasculine	Injection	40–50 mg/week. Paused between month 8 and 9	Between month 6 and 7
P5	24	Nonbinary	Injection	25–50 mg/week	Between month 6 and 7
P6	24	Nonbinary	Injection	60 mg/week	Between month 1 and 2
P7	19	Nonbinary	Gel	12.5–40.5 mg/day	Between month 4 and 5
P8	24	Transgender man	Injection	40 mg/week	Between month 3 and 4
P9	19	Male	Injection	20–40 mg/week	Between month 8 and 9
P10	22	Nonbinary	Injection	30–40 mg/week	Between month 3 and 4
P11	18	Nonbinary	Injection	40 mg/week	Between month 1 and 2

Acoustic Recordings

Participants were asked to complete acoustic recordings in a sound-treated room with an omnidirectional headset microphone (MX153; Shure, Niles, IL). The participant was 7 cm away from the microphone, which was at a 45-degree angle from the midline. Signals were preamplified by an RME Quadmic II (RME, Haimhausen, Germany) and sampled at 44,100 Hz with 16-bit resolution using a MOTI UltraLite-mk3 Hybrid (model UltraLite3Hy; MOTU, Cambridge, MA). The speech signal was recorded using SONAR software (Cakewalk, Boston, MA). Each session, participants were asked to read aloud a standardized reading passage, the Rainbow Passage (Fairbanks, 1960), in their typical speaking voice.

Laryngoscopy

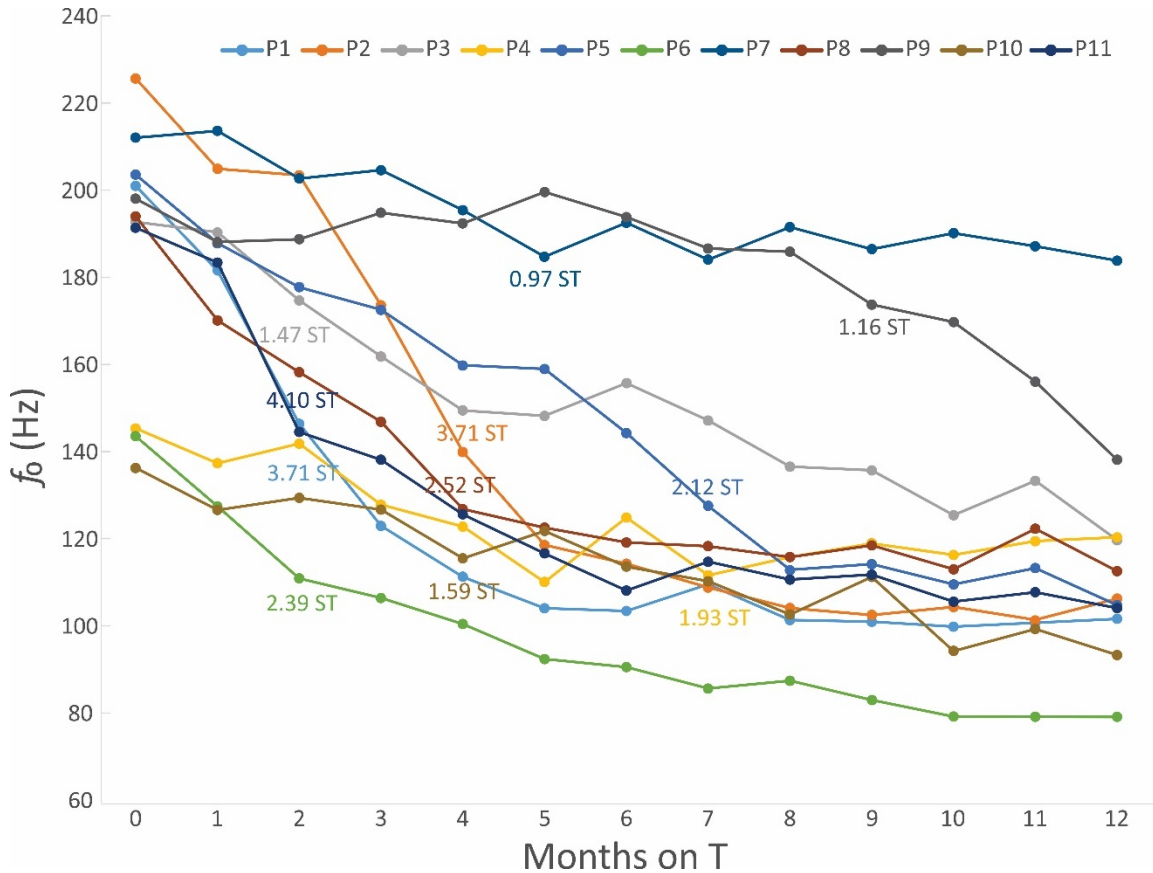
Laryngoscopic exams were collected during the baseline visit and at months 1, 2, 3, 5, 7, 9, 11, and 12 post-T. They were conducted with a flexible trans-nasal laryngoscope with a Digital Stroboscopy System (Kay Elemetrics, Lincoln Park, NJ) with strobe light source via a distal imaging chip (light source and video processor EPK-1000; pediatric endoscope, VNL-1070STK, 3.3 mm width; both Pentax, Tokyo, Japan). Video was digitized at 30 frames/second with a frame size of 480 x 360 pixels. Participants were asked to complete speech tasks during the endoscopic exam, which included vocal productions of sustained /i/ at a comfortable pitch and loudness. Participants were also asked to inhale through their nose to elicit abduction of the vocal folds.

Data Analysis

Acoustic data

The acoustic recording of the Rainbow Passage was used to measure average f_0 in continuous speech. Average f_0 was measured using Praat (Boersma & Weenink, 2015). Participants' audio recordings were manually tracked by a trained researcher to crop out instances of non-speech sounds (e.g., a laugh, cough, etc.), and adjust pitch settings to avoid pitch mistracking errors. Mean f_0 of the whole recording was extracted, and the difference between f_0 for each visit was measured in semitones (ST) in order to determine at which visit the largest drop in f_0 occurred. We only evaluated the first 9 months of treatment, as we expected a plateau in f_0 later on in the first year of T (Figure 1). One participant (P9) experienced a large drop in f_0 following this cutoff, at month 12 (see Figure 1). The trained researcher repeated 15% of samples from a reliability set to calculate intra-rater reliability. An additional trained researcher repeated 15% of the samples to calculate inter-rater reliability. If there was little to no acoustic change (less than 2 ST) in a participant's voice over the course of one year, or less than 0.75 ST between months prior to month 9, they were excluded from the study, as constriction as a result of hyperfunction was not likely to change if there were no structural changes.

Figure 1. Rainbow Passage fundamental frequency (f_0) for TGNC individuals taking T in year 1. Text labels were added to identify at which month each participant's largest drop in f_0 occurred and to quantify the f_0 change in semitones (STs) to visualize variation in acoustic changes. Each color represents an individual participant.



Laryngoscopic data

The laryngoscopic data were used to measure changes in supraglottic constriction as an indicator of new hyperfunctional voicing patterns. Supraglottic constriction was measured using the laryngoscopic exam from the baseline visit, month one, and the month immediately following the participant's largest monthly decrease in f_0 . If there was no laryngoscopic exam collected following the largest monthly decrease in f_0 , then the next monthly visit that included a laryngoscopic exam following the decrease was

analyzed (month 6 for P2; month 5 for P8 and P10). A trained researcher prepared to analyze participant's laryngoscopic exams by manually cropping 2-5 second clips from a full laryngoscopic exam. These clips included the participant producing a sustained modal /i/ and a clip during inspiration (vocal folds being fully abducted). The majority of the clips during inspiration were taken as the participant was cued to sniff in. If this clip was not usable due to poor video quality or obstructed view, we chose a clip where the participant was preparing for phonation. These videos were used to measure supraglottic constriction via VALI ratings and the replicated quantitative measures (Behrman et al., 2003; Han et al., 2023; Ogawa et al., 2013; Poburka et al., 2017; Stager et al., 2001). Quantitative measures of supraglottic constriction were taken from a baseline laryngoscopic exam collected before participants start T, during the first visit after starting T as a presumed control, and at a third timepoint, determined by when their mean f_0 decreased the most from the previous visit, measured in ST.

VALI

Five licensed voice-specialized Speech-Language Pathologists (SLPs) used a custom MATLAB script to view and evaluate the cropped video clips of modal /i/ for each participant at each time point. The SLPs were instructed to provide judgements of supraglottic activity on both the AP and ML planes using an adapted version of the VALI supraglottic activity parameter. Supraglottic constriction was rated on a scale from 0 (no constriction) to 5 (max constriction; Poburka et al., 2017). Each SLP also screened each video for atypical findings. There were no atypical findings within this dataset (i.e., phonotrauma, granuloma, nonvibratory observations). Each SLP repeated 15% of

samples to calculate intra- and inter-rater reliability.

Quantitative Measures of Supraglottic Activity

Custom MATLAB scripts were designed to replicate three of the four quantitative measures analyzed as described in the literature (Behrman et al., 2003; Han et al., 2023; Ogawa et al., 2013). A trained researcher used the three custom scripts to quantify AP and ML constriction for each video of the sustained modal /i/ and rest using the laryngeal outlet, vocal fold width and length, and false vocal fold curvature. Behrman and colleagues (Behrman AP and Behrman ML) used sustained modal /i/ to measure vocal fold width, vocal fold length, and the laryngeal outlet (Behrman et al., 2003). Han and colleagues (Han Abducted ML and Han Adducted ML) used phonation of /i/ and maximum inspiration/end expiration of quiet breathing to measure false vocal fold curvature (Han et al., 2023). Ogawa and colleagues (Ogawa AP and Ogawa ML) used sustained phonation of /i/ and the maximal opening of the vocal folds during inspiration to measure the laryngeal outlet (Ogawa et al., 2013).

Three unique timepoints were chosen for analysis of quantitative measures. These included the baseline visit, month 1 on T, and the month following the largest drop in f_0 . Each timepoint included six quantitative measure analyses. Each participant had 18 outcome measures of supraglottic constriction total. There were two VALI-derived measures per monthly visit (AP and ML constriction for modal /i/). There were two categories of quantitative outcome measures, AP constriction and ML constriction. Outcome measures of AP constriction consisted of Behrman AP and Ogawa AP. Outcome measures of ML constriction consisted of Behrman ML, Han Abducted ML,

Han Adducted ML, and Ogawa ML. The timepoints were chosen to observe differences alongside acoustic changes in the voice on GAHT. Therefore, the third timepoint corresponded to the point at which each participant's largest drop in f_0 occurred, prior to the 9th month on T, as we expected the voice to not change drastically between months 9 and 12 (Cler et al., 2020; Damrose, 2009; Deuster et al., 2016; Irwig et al., 2017; van Borsel et al., 2000).

Statistical Analysis

We hypothesized that existing quantitative metrics of AP and ML constriction would be related to the VALI metrics of AP and ML supraglottic constriction. To test this hypothesis, we compared quantitative measures and a visual-perceptual measure of supraglottic constriction to determine if quantitative measures predicted visual-perceptual measure outcomes. All statistical analyses were completed in R statistical software (RStudio Team, 2020). Significance for all statistical testing was set *a priori* at $\alpha = 0.1$, due to small sample size. We constructed two linear mixed-effects models using the **lmer** package in RStudio, one analyzing all AP measures of supraglottic constriction, and one analyzing all ML measures of supraglottic constriction. We used these models to determine if independent variables (quantitative measures of AP and ML constriction) would predict the dependent variable (visual-perceptual measure of AP ML constriction). Repeated measures were accounted for in the model; timepoint and participant were added to the model as random effects. The model specifications were the following for AP and ML models, respectively: Avg VALI AP ~ Behrman AP + Ogawa AP + (1|Timepoint) + 1|PtID Thesis); and Avg VALI ML ~ Behrman ML + Han Abducted ML

+ Han Adducted ML + Ogawa ML + (1|Timepoint) + (1|PtID Thesis). For significant effects, we calculated β values from the model.

We also hypothesized that the measure designed by Behrman and colleagues would be most closely related to the VALI, because it measures the AP and ML constriction levels compared to the laryngeal outlet most similarly to VALI ratings (Behrman et al., 2003). To test this hypothesis and determine whether the measure designed by Behrman and colleagues is more closely related to the VALI, we determined statistical significance of each independent variable within our linear regressions, and interpreted effect sizes for statistically significant variables.

Results

Outcome measures as a function of time

Outcome measures were analyzed and plotted for visualization. Plots were grouped together by type of outcome measure (visual-perceptual vs. quantitative). Quantitative measure plots were grouped together by direction of constriction (AP vs. ML). There were two visual-perceptual outcome measures (VALI AP and VALI ML) and six quantitative outcome measures, which were grouped as AP (Behrman AP and Ogawa AP) and ML (Behrman ML, Han Abducted ML, Han Adducted ML, and Ogawa ML).

Figure 2. Visual-perceptual (VALI) measures of supraglottic constriction (AP and ML). Baseline was defined as month 0 on T, Control was defined as month 1 on T, and Largest Drop was defined as the month at which the participant’s largest monthly drop in f_0 occurred, prior to month 10. As the values on the y-axis increase, this corresponds with “more constriction.” These parameters also apply to Figures 3 and 4.

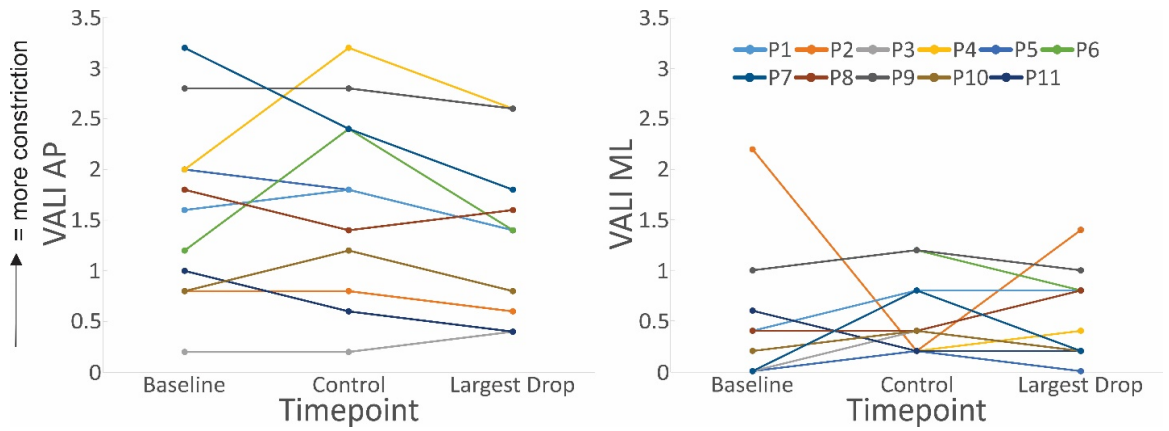


Figure 3. Two quantitative measures of AP supraglottic constriction (Behrman AP and Ogawa AP). Ogawa and colleagues (2013) defined “more constriction” as corresponding to lower values, so the y-axis for these measures is flipped for comparison across measures.

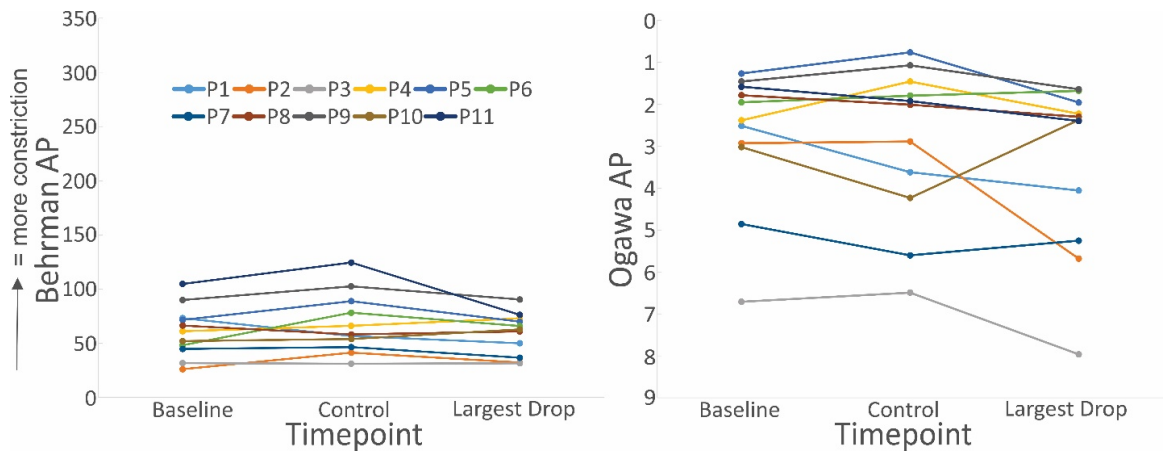
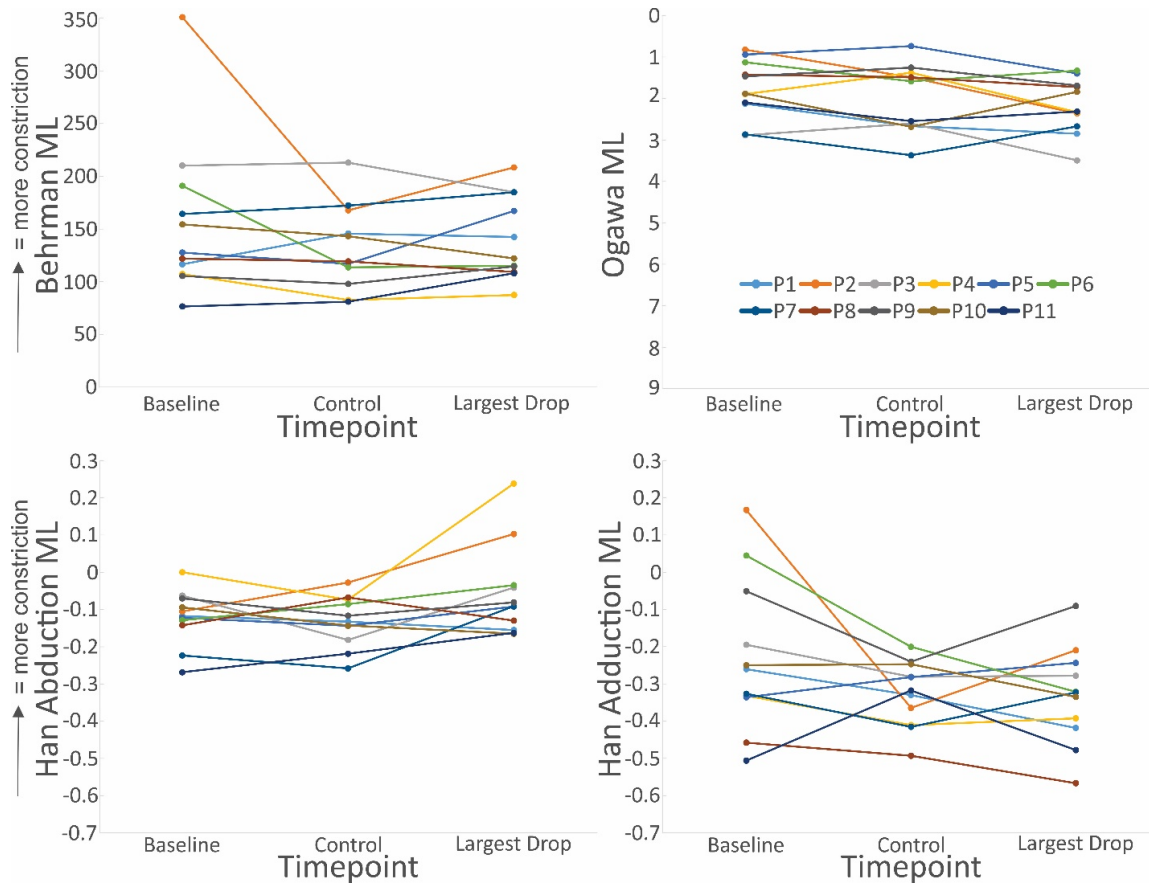


Figure 4. Four quantitative measures of ML supraglottic constriction (Behrman ML, Ogawa ML, Han Abducted ML, and Han Adducted ML). Ogawa and colleagues (2013) defined “more constriction” as corresponding to lower values, so the y-axis for these measures is flipped for comparison across measures.



Reliability

Reliability was calculated for f_0 data, VALI data, and quantitative outcome measure data. For f_0 , Pearson’s r correlations were used to calculate reliability. Inter-rater reliability and intra-rater reliability for f_0 were both $R^2 = 0.99$. For the VALI, average fixed rater intraclass correlation coefficients (ICCs) were used to calculate reliability. For the quantitative outcome measures, single fixed rater ICCs were used to calculate reliability. Inter-rater reliability for VALI AP was 0.92. Inter-rater reliability for VALI ML was 0.81. Reliability for VALI in the AP direction was excellent, whereas in the ML

direction, it was good (Koo & Li, 2016). Inter-rater reliability for Behrman AP and Ogawa AP was excellent (ICC = 1.0 for both measures). Inter-rater reliability for Behrman ML and Ogawa ML was excellent (ICC = 0.98 and 0.97 respectively); Han Abducted ML and Han Adducted ML was moderate (ICC = 0.75 and 0.71 respectively). Intra-rater reliability for Behrman AP and Ogawa AP was excellent (ICC = 0.99 and 1.0 respectively). Intra-rater reliability for Behrman ML, Han Abducted ML, and Ogawa ML was excellent (ICC = 0.98, 0.97, and 0.97 respectively); Han Adducted ML was moderate (ICC = 0.58; Koo & Li, 2016)

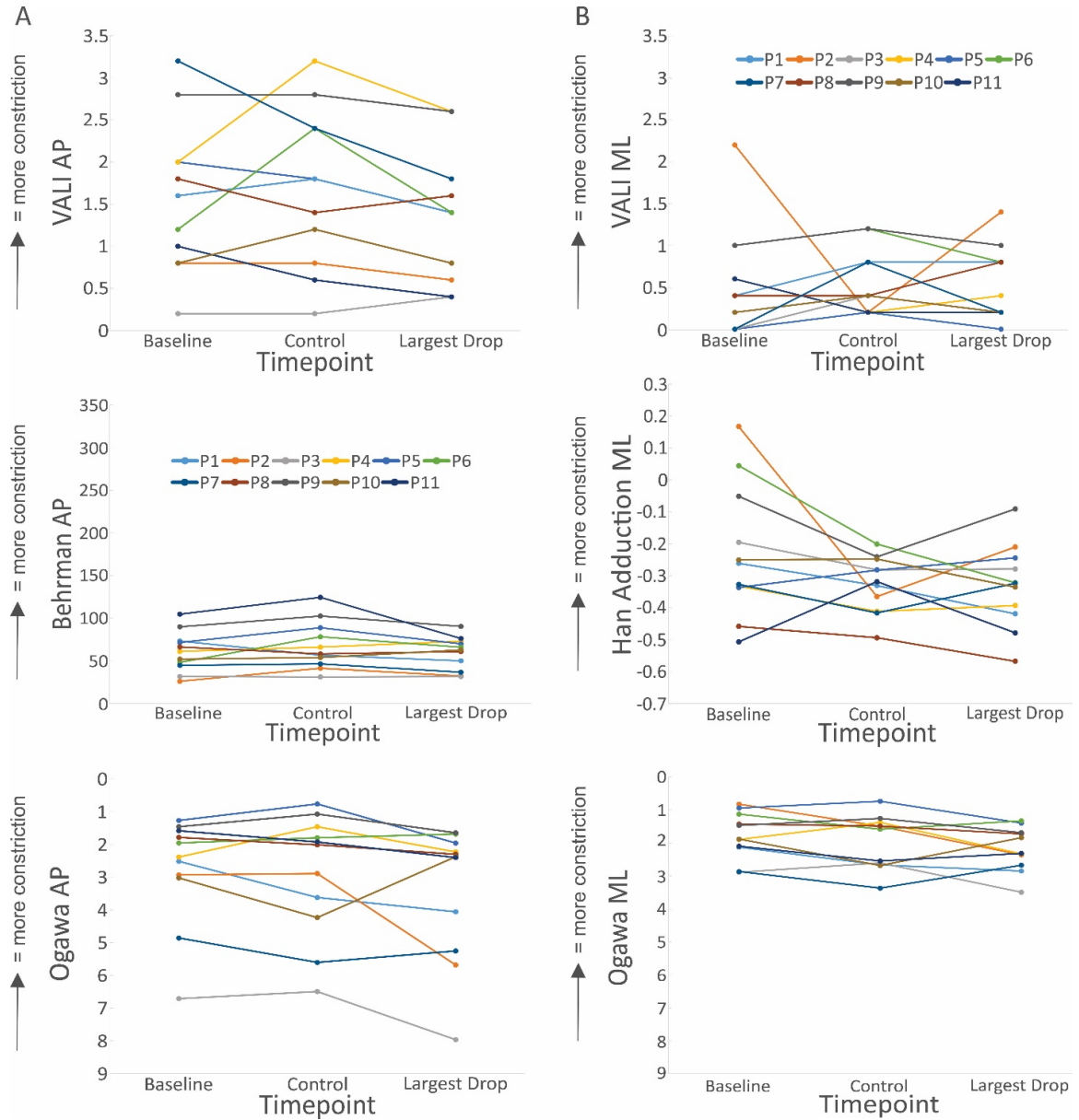
Comparing the VALI to quantitative measures

Statistical analysis was completed to compare the VALI measures (visual-perceptual measures) to quantitative measures of supraglottic constriction. Two linear mixed-effects models were constructed for the AP and ML directions, with the quantitative measures of constriction as the independent variables, and the VALI as the dependent variable. Participant and timepoint were entered as random effects. VALI AP was compared to quantitative measures of AP constriction (Behrman AP and Ogawa AP), and VALI ML was compared to quantitative measures of ML constriction (Behrman ML, Han Abducted ML, Han Adducted ML, and Ogawa ML).

In the AP direction, both Behrman AP and Ogawa AP were found to significantly predict VALI AP outcomes ($p < 0.001$ and $p < 0.1$, respectively). The effect size of Behrman AP was .02 and the effect size of Ogawa AP was -0.12. In a correlation of fixed effects, Behrman AP and Ogawa AP were relatively correlated, at $r = 0.621$. In the ML direction, Han Adducted ML and Ogawa ML were found to be statistically significant,

indicating that they predict the VALI ML outcomes ($p < 0.01$; $p < 0.05$ respectively), whereas Behrman ML and Han Abducted ML did not predict VALI ML outcomes. The effect size of Han Adducted ML and Ogawa ML were 1.72 and -0.21, respectively. In a correlation of fixed effects, Behrman ML and Han Adducted ML were relatively correlated, at $r = -0.657$. We ran variable inflation factors for all fixed effects and values were all < 10 , suggesting that multicollinearity was not a major concern.

Figure 5. On the left column (A), the two AP measures of constriction which predict the VALI measures are represented. On the right column (B), the two out of four ML measures of constriction which predict the VALI measures are represented.



Discussion

Gender-affirming voice therapy is a necessity for many individuals. Although it is most often the case that individuals receiving this care are assigned male at birth (AMAB) and seeking feminization, it may not be so one-sided. Specifically, there has been self-report of TGNC AFAB individuals incurring voice problems (Azul et al., 2018; Cler et al., 2020; Nygren et al., 2016). This necessitates an investigation of hyperfunctional symptoms throughout the process of GAHT use. The purpose of this study was to determine which quantitative measures of supraglottic constriction were most closely related to visual-perceptual measures of supraglottic activity in TGNC people who are using T as a form of GAHT. Specifically, we wanted to explore whether instances of vocal hyperfunction, as measured by supraglottic constriction, occur alongside acoustic changes in f_0 , which may be associated with structural laryngeal changes. Large and fast changes in f_0 may imply large structural changes, which may lead to supraglottic constriction. This is due to adaptation needs in the supraglottic area, which could induce constriction, eventually leading to voice problems. We found that there was variation in supraglottic constriction across participants over the course of a year on T, and that this variation did not obviously correlate with periods of large structural changes, as was expected at the largest monthly decrease in f_0 . There was variation across the range for the visual-perceptual and quantitative measures of AP constriction, whereas we saw little variation across the range for measures of ML constriction (see Figures 2, 3, and 4).

Over the course of one year on T, 7 out of 11 participants had their largest monthly decrease in f_0 occur within 4 months of starting treatment. Another participant experienced their largest monthly decrease in f_0 between months 4 and 5. This finding corroborates the findings of previous studies investigating voice changes in this population (Cler et al., 2020; Damrose, 2009; Deuster et al., 2016; Irwig et al., 2017; van Borsel et al., 2000). Although the majority of participants experienced their largest monthly decrease in f_0 within 6 months of starting treatment, several participants experienced decreases later on (between months 7 to 9; see Figure 1; P4, P5, and P9), and some participants experienced multiple timepoints with large decreases (see Figure 1; P1 and P2) as well as plateaus (where their f_0 stabilized after about month 6) throughout their first year on T. This may explain why our findings with regard to supraglottic constriction were more variable, as change in f_0 over time was not linear or consistent. If supraglottic constriction increased at timepoints of large changes in f_0 , we would have observed an upward trend in the line graphs at the largest drop timepoint, following a stagnation in the trend from baseline to control (see Figures 2, 3, and 4). Therefore, structural changes leading to functional changes (i.e., changes in constriction) were also less likely to be linear. Additionally, functional voicing pattern changes may take more time for individuals to adapt to, meaning that changes in hyperfunctional effects (i.e., supraglottic constriction) may only occur following long periods of varying functional voicing pattern. Functional changes may also depend on where a participant started regarding their baseline f_0 (see Figure 1 for variation in participants' baseline f_0).

We hypothesized that quantitative measures of supraglottic constriction would predict visual-perceptual measures. This hypothesis was partially supported by results of this study. In the AP direction, the results revealed that both of the quantitative measures examined did predict visual-perceptual measures from the VALI, whereas in the ML direction, only two out of four (Han Adducted ML and Ogawa ML) of the quantitative measures predicted VALI outcomes. It was also hypothesized that the measure designed by Behrman and colleagues would most closely predict the VALI outcomes. This measure did more closely predict VALI outcomes, but only in the AP direction. This measure did not significantly predict VALI outcomes in the ML direction. Both directions seem to have similar results to VALI outcomes, as each direction had two measures that were significant. The ML direction had fewer measures that were significant (2 of 4 versus 2 of 2), so in the future clinicians and researchers should use caution in implementing the two metrics that were not significantly related to the VALI.

AP direction

In the AP direction, both Behrman and Ogawa significantly predicted VALI outcomes ($p < 0.001$ and $p < 0.1$, respectively). When evaluating effect size using standardized β values, the model showed that both quantitative measures had small effects (Nieminen, 2022). Ogawa AP had a slightly larger effect at -0.12 compared to Behrman AP at 0.02. Ogawa is expected to have a negative effect when evaluating effect size, as the authors of that measure define “more constriction” as having a smaller value, whereas the Behrman and VALI measures both defined “more constriction” as having a larger value (Behrman et al., 2003; Ogawa et al., 2013; Poburka et al., 2017). These

findings may indicate that quantitative measures of AP constriction are predictors of visual-perceptual measures, as both measures were significant, but that they only explain a small part of the variance in VALI scores ($\beta = -0.12$ for Ogawa AP and $\beta = 0.02$ for Behrman AP). Of note, correlation of fixed effects between Behrman AP and Ogawa AP was $r = 0.621$ (moderate-strong), as observed in Figure 5. This suggests that these two measures may be quantified in similar ways. These findings suggest that AP quantitative measures are good predictors of visual-perceptual measures, as they are measuring similarly and both predicting VALI outcomes.

Given the results of the AP direction linear mixed-effects model, we can interpret that Behrman and Ogawa may measure AP supraglottic constriction in similar ways. However, these two measures also have differences in their measurement techniques. Behrman and Ogawa both take the length and width measurements of the visible true vocal folds, and normalize those values to a constant. Behrman normalizes using the laryngeal outlet of the same videostroboscopic image (dividing the length and width of the vocal folds by the laryngeal outlet to determine AP and ML constriction), whereas Ogawa normalizes using the mean length of the vocal folds at inspiration of a different videostroboscopic image (dividing the length and width of the vocal folds by the mean length of the vocal folds at inspiration to determine AP and ML constriction; Behrman et al., 2003; Ogawa et al., 2013). These differences in normalization are both potentially problematic, as normalizing to a value that is not standard does not effectively standardize a measurement, due to the changing nature of the value being used for normalization. However, despite these potential problems, their prediction of VALI

outcomes (the current research standard) demonstrates some confidence in their ability to capture changes in supraglottic activity.

ML direction

In the ML direction, Han Adducted and Ogawa were significant ($p < 0.01$ and $p < 0.05$, respectively) predictors of VALI ML outcomes. However, Behrman ML and Han Abducted ML were not shown to significantly predict VALI outcomes. In the case of Han Abducted, this may be due to using images which were not used by the VALI (abduction images). Surprisingly, the Behrman also did not predict VALI outcomes, even though the use of the laryngeal outlet to determine ML (and AP) constriction is similar to the perceptual ratings of the VALI. This may be due to the quantitative measure being more sensitive to change than the perceptual nature of the VALI. The quantitative measure uses measurements of pixels to determine percentage of supraglottic constriction, whereas the VALI relies on perception alone to complete a rating (Behrman et al., 2003; Poburka et al., 2017). The standardized β values for Ogawa ML and Han Adducted ML were -0.21 (small) and 1.72 (large), respectively (Nieminen, 2022). Han Adducted ML experienced a much larger effect on VALI outcomes than Ogawa ML, explaining much more of the variance of the VALI ML ratings. As mentioned previously, the Ogawa measure's description of which values indicate more constriction are opposite that of all other measures, which gives explanation for the negative value. Of note, in the correlation of fixed effects, Behrman ML and Han Adducted ML were moderately-strongly negatively correlated, with a value of $r = -0.657$. This may indicate that despite Behrman ML not being significant in the model, its high correlation with a significant

measure could be interpreted as indicating significance that was masked in the model. Due to this potential outcome, we ran variable inflation factors for all fixed effects and values were all <10 , suggesting that multicollinearity was not a major concern.

Two quantitative measures of ML constriction significantly predict VALI outcomes, and one in particular (i.e., Han Adducted ML) had a larger effect on VALI outcomes. There were also two quantitative measures of ML constriction which did not significantly predict VALI outcomes. These findings support the use of Han Adducted ML and Ogawa ML (the two significant measures) as better tools to use clinically, whereas Behrman ML and Han Abducted ML (the two measures which were not significant) should be used with caution. With the exception of intra-rater reliability for Han Abducted ML, both inter- and intra- rater reliability for both Han measures were moderate, whereas all other visual-perceptual and quantitative measures had excellent reliability. These findings may indicate that although Han Adducted ML significantly predicted VALI ML outcomes, the agreement between raters was less reliable, meaning that these outcomes may not be replicated in future use.

We found that there was little variation across the range for ML ratings on the VALI, as compared to the ratings for VALI AP, which used a greater range of the scale (see Figure 2). Some quantitative measures were significant, while others were not. These findings bring into consideration whether ratings of the VALI in the ML direction can be made in confidence, as the quantitative measures did not always accurately predict the visual-perceptual measure. The range of ratings for VALI ML was very small (most ratings either 0 or 1 out of a scale of 5 with 0 being no constriction and 5 being max

constriction, with the exception of P2; see Figure 2). This small range may have either been due to the scale of the VALI being unable to capture the variance in ML constriction, or there not being much ML constriction present. The latter having been predicted by previous studies showing skepticism around the measure as it relates to vocal hyperfunction (Dabirmoghaddam et al., 2021; Shembel et al., 2023). Poburka and colleagues (2017) discussed that in their study using the VALI, supraglottic activity, specifically in the ML direction, had the highest instances of rating difficulty. Thus, there may be difficulty overall with supraglottic constriction ratings, and in general, visual-perceptual measures may not be the best for understanding true values of constriction in different populations. In future studies, there may need to be a different or smaller binary ML scale to mark presence or absence, as the large scale and rater's understanding of constriction may be merging to cause challenges in our understanding of ML constriction overall.

Clinical implications

Many TGNC individuals identify gender-affirming voice therapy as a necessity for their gender transition. This research allows us to understand more about how supraglottic constriction as a measure of vocal hyperfunction might be changing across TGNC individuals taking T. The eleven participants we have included in this study had a large range in acoustic variation across the longitudinal timeline, but all experienced a drop of at least 2 ST in their average f_0 during a reading passage. Three participants were excluded from the study due not experiencing equal to or greater than a 2 ST drop in f_0 across 1 year on T, or not experiencing a drop greater than 0.75 ST prior to month 9 on T.

All of this variation that occurs requires individuals to use their voice in different ways, which may lead to voice difficulties such as fatigue and voice quality/control challenges (Azul et al., 2018; Cler et al., 2020; Nygren et al., 2016).

We found that supraglottic constriction did not change in the expected direction between baseline and the largest decrease in f_0 , but rather that there was variation in supraglottic constriction across timepoints and participants. There may be changes in structure and function of the voice during the course of a year on T, but not at a naturally expected time, based on when large and fast structural changes are occurring and when individuals in previous studies have noted hyperfunctional vocal qualities that they have incurred.

Our findings suggest that there may be an individualized need for addressing constriction changes, but only if an individual also presents with voice symptoms. More individuals may need voice therapy during the first year of taking T. Even though T is thought to and has shown drastic decreases in f_0 that align with that of cisgender men having gone through natal T driven puberty, this was not the case for all participants of our study (Hodges-Simeon et al., 2021). The presence of variable supraglottic constriction in these individuals may necessitate more behavioral intervention when paired with self-report of voice symptoms. Structural changes due to T are often thought to do all of the work to achieve a voice that better aligns with the individual's gender identity, but this may not always be the case (Azul et al., 2017). Perception of voice symptoms was not assessed in this study, but it should be the most important consideration when providing care to this population. Although there is currently a

standard way of measuring constriction in research (visual-perceptual, the VALI), there is no current standard clinical measure for supraglottic constriction. Clinically, an otolaryngologist (ENT) informally evaluates the videostroboscopic exam and perceptually observes functional voicing patterns of the glottis and structures above it. How we measure constriction is not always the most feasible or most sensitive to the changes that are occurring, especially in a clinical setting. However, the VALI was rated using videos, whereas the quantitative analysis was done using images. It may be more ecologically valid to watch a video and make judgements, which is what is done in a clinical setting. This also means that an “average” amount of constriction needs to be determined, whereas the quantitative measures utilized one time point.

In this case, we found that different tools may be best for different directions (AP vs. ML), but that the Ogawa measure was significant (meaning it might predict VALI outcomes) in both directions. In the AP direction, the Behrman measure may also predict VALI outcomes, where in the ML direction the Han measure in the Adducted condition may predict VALI outcomes. These measures that were found to be significant and potentially predict VALI outcomes may be the best tools to use clinically when wanting more precise measurements of potential hyperfunction or determining best practice and individualized care. Specifically, ratings and measures of AP constriction may be more useful in determining functional effects on the voice when compared to ML ratings and measures

Limitations and future directions

This study had several limitations that should be noted. Constriction in both quantitative and visual-perceptual measures was measured during endoscopy and not in a naturalistic conversational environment. Although this is helpful in understanding more about the structure and function of the voice while undergoing GAHT with T, it does not explain what may be happening when individuals are using their voice in conversation for extended periods of time. The flexible endoscope also added an extra factor to the participant's voicing outcomes. The endoscope passing through their nose may have caused them to have more tension or hyperfunction while voicing that may not be there under typical conditions. Using an endoscope is necessary to diagnose supraglottic constriction (which may be indicative of hyperfunction) in individuals, but the discomfort of the scope still may be a factor in participants with no related voice disorders (Koufman & Blalock, 1991 as cited in Behrman et al., 2003). We also did not compare measurable constriction to self-reported perception of voice symptoms, so constriction that is visualized may not be indicative of hyperfunction in these individuals. Lastly, we had a small sample size, so results should be interpreted with caution.

It is also important to note that the measures used for objective analysis were quantitative in nature, as they utilized numbers in pixels to determine amount of constriction, but similarly to VALI ratings, they can still be prone to human error. For the quantitative measures, there were guidelines surrounding where images should be marked and anatomical locations utilized for placement of markers, but determination of specific locations was not automated and rather done based on perception. This was a manual

process which involved selecting images and making markings based on what was perceived to be the area that the guidelines pointed to. On a related note, two of the quantitative measures were normalized in different ways to determine amount of constriction in both directions. To have a more accurate depiction of change across timepoints and images, normalization should be made to something that is constant, as lengths are sensitive to how far away the endoscope is from the larynx.

Conclusion

In this study, several (but not all) quantitative measures analyzed predicted the VALI outcomes (all AP measures, and only two out of four ML measures) in TGNC individuals taking T for the first time. For this study sample, the VALI results showed variation across the full range of the scale in the AP direction, whereas there was little variation in the ML direction. These results suggest that AP measures can be used to determine variation in constriction clinically, but that ML measures should be used with caution. The results also help to explain more about what is occurring in the larynx during large and fast changes to its structure in individuals who were AFAB and take GAHT with T. Though we are able to see variation in some measures of constriction, the constriction did not increase at the timepoint we expected (following a large decrease in f_0). Rather, there was variation in constriction across the year following the start of T. While these findings are an important step forward in research related to potential changes in function in these individuals, more work is required to compare this variation in constriction measures with self-reported symptoms, and if necessary, individualize evaluation and treatment of voice in AFAB people on T.

BIBLIOGRAPHY

- ASHA. (2005). *Guidelines for Manual Pure-Tone Threshold Audiometry* [Online post]. American Speech-Language-Hearing Association. <https://www.asha.org/policy/gl2005-00014/>
- Azul, D., Arnold, A., & Neuschaefer, -Rube Christiane. (2018). Do Transmasculine Speakers Present With Gender-Related Voice Problems? Insights From a Participant-Centered Mixed-Methods Study. *Journal of Speech, Language, and Hearing Research, 61*(1), 25–39. https://doi.org/10.1044/2017_JSLHR-S-16-0410
- Azul, D., Nygren, U., Södersten, M., & Neuschaefer-Rube, C. (2017). Transmasculine People’s Voice Function: A Review of the Currently Available Evidence. *Journal of Voice, 31*(2), 261.e9-261.e23. <https://doi.org/10.1016/j.jvoice.2016.05.005>
- Behrman, A., Dahl, L. D., Abramson, A. L., & Schutte, H. K. (2003). Anterior-Posterior and medial compression of the supraglottis: Signs of nonorganic dysphonia or normal postures? *Journal of Voice, 17*(3), 403–410. [https://doi.org/10.1067/S0892-1997\(03\)00018-3](https://doi.org/10.1067/S0892-1997(03)00018-3)
- Boersma, P., & Weenink, D. (2015). *Praat: Doing phonetics by computer* [Computer software]. <http://www.praat.org>
- Cler, G. J., McKenna, V. S., Dahl, K. L., & Stepp, C. E. (2020). Longitudinal Case Study of Transgender Voice Changes Under Testosterone Hormone Therapy. *Journal of Voice, 34*(5), 748–762. <https://doi.org/10.1016/j.jvoice.2019.03.006>
- Dabirmoghaddam, P., Aghajanzadeh, M., Erfanian, R., Aghazadeh, K., Sohrabpour, S., Firouszifar, M., Maroufizadeh, S., & Nikraves, M. (2021). Comparative Study of Increased Supraglottic Activity in Normal Individuals and those with Muscle Tension Dysphonia (MTD). *Journal of Voice, 35*(4), 554–558. <https://doi.org/10.1016/j.jvoice.2019.12.003>
- Damrose, E. J. (2009). Quantifying the impact of androgen therapy on the female larynx. *Auris Nasus Larynx, 36*(1), 110–112. <https://doi.org/10.1016/j.anl.2008.03.002>
- Deuster, D., Matulat, P., Knief, A., Zitzmann, M., Rosslau, K., Szukaj, M., am Zehnhoff-Dinnesen, A., & Schmidt, C.-M. (2016). Voice deepening under testosterone treatment in female-to-male gender dysphoric individuals. *European Archives of Oto-Rhino-Laryngology, 273*(4), 959–965. <https://doi.org/10.1007/s00405-015-3846-8>
- Fairbanks, G. (1960). *Voice and articulation drillbook* (2nd ed.). Harper & Row.

- Fernandez, S., Garaycochea, O., Martinez-Arellano, A., & Alcalde, J. (2020). Does More Compression Mean More Pressure? A New Classification for Muscle Tension Dysphonia. *Journal of Speech, Language, and Hearing Research*, *63*(7), 2177–2184. https://doi.org/10.1044/2020_JSLHR-20-00042
- Ferrand, C. T. (2011). *Voice Disorders: Scope of Theory and Practice* (2nd ed.). Pearson Higher Ed.
- Han, J., Morrison, R., Mau, T., & Shembel, A. C. (2023). Quantification of False Vocal Fold Hyperfunction During Quiet Breathing in Muscle Tension Dysphonia. *The Laryngoscope*, *133*(12), 3449–3454. <https://doi.org/10.1002/lary.30814>
- Hancock, A. B., Childs, K. D., & Irwig, M. S. (2017). Trans Male Voice in the First Year of Testosterone Therapy: Make No Assumptions. *Journal of Speech, Language, and Hearing Research*, *60*(9), 2472–2482. https://doi.org/10.1044/2017_JSLHR-S-16-0320
- Hillman, R. E., Stepp, C. E., Van Stan, J. H., Zañartu, M., & Mehta, D. D. (2020). An Updated Theoretical Framework for Vocal Hyperfunction. *American Journal of Speech-Language Pathology*, *29*(4), 2254–2260. https://doi.org/10.1044/2020_AJSLP-20-00104
- Hodges-Simeon, C. R., Grail, G. P. O., Albert, G., Groll, M. D., Stepp, C. E., Carre, J. M., & Arnocky, S. A. (2021). Testosterone therapy masculinizes speech and gender presentation in transgender men. *Scientific Reports*, *11*(3494).
- Irwig, M. S., Childs, K., & Hancock, A. B. (2017). Effects of testosterone on the transgender male voice. *Andrology*, *5*(1), 107–112. <https://doi.org/10.1111/andr.12278>
- Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155–163.
- Koufman, J. A., & Blalock, P. D. (1991). Functional Voice Disorders. *The Otolaryngologic Clinics of North America*, *24*(5), 1059–1073.
- Nieminen, P. (2022). Application of Standardized Regression Coefficient in Meta-Analysis. *BioMedInformatics*, *2*(3), 434–458.
- Nygren, U., Nordenskjöld, A., Arver, S., & Södersten, M. (2016). Effects on Voice Fundamental Frequency and Satisfaction with Voice in Trans Men during Testosterone Treatment—A Longitudinal Study. *Journal of Voice*, *30*(6), 766.e23-766.e34. <https://doi.org/10.1016/j.jvoice.2015.10.016>

- Oates, J., & Winkworth, A. (2008). Current knowledge, controversies and future directions in hyperfunctional voice disorders. *International Journal of Speech-Language Pathology, 10*(4), 267–277. <https://doi.org/10.1080/17549500802140153>
- Ogawa, M., Hosokawa, K., Yoshida, M., Yoshii, T., Shiromoto, O., & Inohara, H. (2013). Immediate Effectiveness of Humming on the Supraglottic Compression in Subjects with Muscle Tension Dysphonia. *Folia Phoniatrica et Logopaedica, 65*(3), 123–128. <https://doi.org/10.1159/000353539>
- Palmer, D., Dietsch, A., & Searl, J. (2012). Endoscopic and Stroboscopic Presentation of the Larynx in Male-to-Female Transsexual Persons. *Journal of Voice, 26*(1), 117–126. <https://doi.org/10.1016/j.jvoice.2010.10.014>
- Patel, R. R., Awan, S. N., Barkmeier-Kraemer, J., Courey, M., Deliyiski, D., Eadie, T., Paul, D., Svec, J. G., & Hillman, R. (2018). Recommended Protocols for Instrumental Assessment of Voice: American Speech-Language-Hearing Association Expert Panel to Develop a Protocol for Instrumental Assessment of Vocal Function. *American Journal of Speech-Language Pathology, 27*, 887–905. https://doi.org/10.1044/2018_AJSLP-17-0009
- Poburka, B. J., Patel, R. R., & Bless, D. M. (2017). Voice-Vibratory Assessment With Laryngeal Imaging (VALI) Form: Reliability of Rating Stroboscopy and High-speed Videoendoscopy. *Journal of Voice, 31*(4), 513.e1-513.e14. <https://doi.org/10.1016/j.jvoice.2016.12.003>
- RStudio Team. (2020). *RStudio: Integrated Development for R* [Computer software]. RStudio, PBC, Boston, MA. <http://www.rstudio.com/>
- Sachdeva, K., & Shrivastava, T. (2018). Dysphonia and its correlation with acoustic voice parameters. *International Journal of Phonosurgery & Laryngology, 8*(1), 6–12.
- Shembel, A. C., Morrison, R. A., McDowell, S., Smeltzer, J. C., Crocker, C., Bell, C., & Mau, T. (2023). *Relationships Between Laryngoscopic Analysis Metrics of Supraglottic Compression and Vocal Effort in Primary Muscle Tension Dysphonia. In Press, Corrected Proof.* <https://doi.org/10.1016/j.jvoice.2023.06.011>
- Solomon, N. P. (2008). Vocal fatigue and its relation to vocal hyperfunction. *International Journal of Speech-Language Pathology, 10*(4), 254–266. <https://doi.org/10.1080/14417040701730990>
- Stager, S. V., Bielowicz, S. A., Regnell, J. R., Gupta, A., & Barkmeier, J. M. (2000). Supraglottic Activity: Evidence of Vocal Hyperfunction or Laryngeal

Articulation? *Journal of Speech, Language, and Hearing Research*, 43(1), 229–238. <https://doi.org/10.1044/jslhr.4301.229>

Stager, S. V., Bielowicz, S., Gupta, A., Marullo, S., Regnell, J. R., & Barkmeier, J. (2001). Quantification of Static and Dynamic Supraglottic Activity. *Journal of Speech, Language, and Hearing Research*, 44(6), 1245–1256. [https://doi.org/10.1044/1092-4388\(2001/097\)](https://doi.org/10.1044/1092-4388(2001/097))

van Borsel, J., de Cuyper, G., Rubens, R., & Destaeke, B. (2000). Voice problems in female-to-male transsexuals. *International Journal of Language & Communication Disorders*, 35(3), 427–442. <https://doi.org/10.1080/136828200410672>

CURRICULUM VITAE

