Effects of Gender-confirming Pitch-raising Surgery in Transgender Women a Long-term Follow-up Study of Acoustic and Patient-reported Data

*Victoria Kelly, †·‡Stellan Hertegård, §Jenny Eriksson, *·§Ulrika Nygren, and *·§Maria Södersten, *†‡§*Stockholm, Sweden*

Summary: Objectives. Transgender women often seek to feminize their voice so that it becomes congruent with their gender identity. Many receive voice therapy (VT) with good results. Some also need pitch-raising surgery. The purpose of this study was to investigate long-term effects and to compare outcomes between cricothyroid approximation (CTA) and glottoplasty (GP).

Methods. This study included retrospective data from 24 patients (35−67 years). Eleven patients had undergone CTA and 13 had undergone GP. Audio recordings were performed in a sound-treated booth and patients answered questionnaires before and after VT, post surgery, and at 1-year follow-up. Fundamental frequency (f_o) measures were extracted from voice range profiles (VRPs) and speech range profiles (SRPs) and were compared with cisgender data.

Results. Minimum f_o in VRP increased significantly only after CTA (95 to 123 Hz). Maximum f_o in VRP was significantly lowered after GP (765 to 652 Hz), even more after CTA (677 to 475 Hz). Speaking mean f_o increased significantly after VT and post surgery, and was thereafter stable (CTA 167 Hz, GP 169 Hz). The maximum f_o in the SRP increased only after VT. The frequency ranges were strongly reduced after CTA. Patients were in general satisfied at follow-up and rated the GP outcomes more favorably than CTA.

Conclusions. Both surgical methods have advantages and disadvantages. The very restricted speaking and physiological frequency ranges, which do not favor an optimal female voice, were mainly found after the irreversible CTA. Thus, CTA is no longer performed at our hospital, whereas surgical techniques based on GP are being developed further. We strongly recommend the use of VRPs to evaluate treatment effects.

Key Words: Voice−Treatment−Transsexual−Transfeminine−MtF.

INTRODUCTION

Transgender individuals experience that their gender identity is not congruent with their body and sex assigned at birth. If they express gender dysphoria, they may, after evaluation by a psychiatric team, receive the diagnosis transsexualism (F64.0), other gender identity disorder (F64.8), or gender identity disorder not otherwise specified (F64.9) according to International Classification of Diseases, Tenth Revision (ICD-10). In Sweden, the national health services offer assessment and gender-confirming medical treatment for transgender people.[1] After confirmation of a gender dysphoria diagnosis, the patient is referred for hormone treatment, voice assessment, germ cell preservation, and plastic surgery consultation. Six psychiatric teams are responsible for diagnostic assessments and coordination of gender-confirming medical interventions in Sweden. The number of applications for a legal sex change has increased significantly over the years,[2] especially after 2013, when a law enacted in 1972 was amended so that sterilization was no longer required.

Many transgender women want to feminize their voices to match their self-identified gender, so they will be perceived as female in communication with others, which may reduce the gender dysphoria. Therefore, interventions to feminize the voice are important parts of gender-confirming treatment, and many transgender women seek voice therapy (VT). Exercises to increase fundamental frequency (f_o) are essential, but it is usually also necessary to adapt other aspects of speech and voice production, such as voice quality, intonation, articulation, resonance, and voice sound pressure level (SPL) (eg, see References 3−5). An increasing number of studies have shown positive effects of VT; see review articles by Oates[6] and Oates and Dacakis.[7] For many patients, VT is enough to find a voice in congruence with their gender identity. Others may also need pitch-raising surgery to raise or maintain female speaking pitch, or to abolish voice production in the lowest frequency range, for example, during coughing and sneezing. Several studies have been conducted evaluating the effects of pitch-raising surgery. However, these studies and studies about VT outcomes often suffer from being based on a small number of participants and lack of control groups.[8]

There are several surgical techniques to raise pitch in transgender women, such as cricothyroid approximation (CTA)[9]; CTA supported with miniplates[10]; Wendler glottoplasty (GP)[11,12]; laser-aided voice adjustment[13]; feminization laryngoplasty, including thyrohyoid approximation,[14] transmuscular anterior suturing, and stiffening of the vocal folds (VFs) with cordotomy incision using CO_2 laser[15]; and Vocal Fold Retrodisplacement advancement of the anterior commisure (VFSRAC).[16] It is difficult to compare results between studies

Accepted for publication March 8, 2018.

Conflict of interest: This is original work and the manuscript has not been submitted to other journals. The authors do not have any conflicts of interest.

From the *Functional Area Speech & Language Pathology, Karolinska University Hospital, Stockholm, Sweden; †Department of Otorhinolaryngology, Karolinska University Hospital, Stockholm, Sweden; ‡Department of Clinical Science, Intervention and Technology, Division of Ear, Nose and Throat Disease, Karolinska Institutet, Stockholm, Sweden; and the §Department of Clinical Science, Intervention and Technology, Division of Speech and Language Pathology, Karolinska Institutet, Stockholm, Sweden.

Address correspondence and reprint requests to Victoria Kelly, Functional Area Speech & Language Pathology, M65, Karolinska University Hospital, Huddinge, SE-141 86 Stockholm, Sweden. E-mail: victoria.kelly@sll.se

Journal of Voice, Vol. 33, No. 5, pp. 781−791

0892-1997

© 2018 The Voice Foundation. Published by Elsevier Inc. All rights reserved.

https://doi.org/10.1016/j.jvoice.2018.03.005

because different outcome measures have been used and evaluations have been conducted at different follow-up times. Short-term results, evaluated less than 1 year after surgery, have shown that CTA raises speaking f_o but also reduces the pitch range.[10,17,18] Few long-term results evaluated more than 1 year after surgery have been published. Neumann and Welzel[10] found no significant changes in speaking pitch 1 year after surgery compared with directly after surgery, showing that the results are stable over time. Negative results such as hoarseness and limited frequency range have been reported after CTA, and laser reduction GP has been used to improve voices after failed CTA.[19] In a recent systematic review of efficacy studies of pitch-raising surgery, 3185 publications were identified, from which 20 were selected for further analyses, based on satisfactory scientific quality.[8] Eight of the studies involved CTA, six used anterior web formation, and six used other types or combinations of surgery. Among the results was a substantial rise in speaking f_o after surgery for all surgery techniques. Dysphonia increased, based on perceptual analysis, mainly after web formation techniques. An even more recent systematic review and meta-analysis of 13 studies confirmed some of the findings by Van Damme et al[8] but concluded that the VF shortening techniques appeared to result in a larger increase of f_o and that more dysphonia was found after CTA.[20] A majority of patients in the different studies were satisfied after surgery. Van Damme et al[8] and Song and Jiang[20] concluded that studies with stronger research design with focus on long-term results after pitch-raising surgery are needed. So far, no studies have compared two surgical methods in a systematic way.

Since 2004, we have worked according to a structured program for transgender patients at Karolinska University Hospital.[21] Transgender women are referred from the psychiatric gender team when a gender incongruence diagnosis has been confirmed. The voice assessment, conducted by a speech language pathologist, includes an interview, recordings of speech range profiles (SRPs) and voice range profiles (VRPs) in a sound-treated booth, and collection of patient-reported data from questionnaires. Laryngoscopy and examination of the thyroid notch are performed by an otorhinolaryngologist who specializes in phoniatrics. If the patient needs gender-confirming feminizing voice treatment to decrease gender and voice dysphoria, VT is the first choice of intervention. Depending on the results, some patients may also need pitch-raising surgery. Criteria for surgery are difficulties to raise or maintain female speaking pitch despite VT or phonation in a very low frequency range during involuntary voice production, for example, during coughing and sneezing. During the period of the study, approximately 20% of the patients were operated on at our hospital after VT.

The purpose of the present study was to evaluate long-term results, at least 1 year after pitch-raising surgery, in transgender women and to compare results between the two surgical methods, CTA and GP. The following questions were addressed: (1) Is the physiological frequency range affected by pitch-raising surgery? If so, how much and has it reached cisfemale range? (2) Is speaking f_o affected by pitch-raising surgery? If so, how much and has it reached cisfemale range? (3) Are the patients satisfied with the voice after surgery? (4) Are there voice-related complications after surgery? (5) What surgical method should be recommended?

METHODS

Data were collected retrospectively from medical records, voice recordings, and patient-reported data from questionnaires. Inclusion criteria were availability of data from before surgery and from follow-up at least 12 months after surgery. For several of the participants, data were also available from their first visit before starting VT, as well as after VT, and 2−3 months postoperatively.

Ethical approval was obtained from the Stockholm Regional Ethical Board (No. 2016/2483-31/2).

Participants

Of 32 patients who underwent surgery after VT, a total of 24 transgender women were included in the study. The eight patients who were not included either declined participation or did not show up at follow-up because they lived far away or were satisfied with the outcomes and wanted to go on with their lives. Of those who declined five patients had surgeries with GP and three had surgeries with CTA. All patients had undergone pitch-elevating surgery at Karolinska University Hospital in Stockholm during the period 2003−2015. All patients were presenting themselves in the self-identified female gender during the period of VT and had had hormonal therapy for a length of time before surgery. Surgery was performed by one surgeon (author SH). The patients' ages at follow-up ranged between 27 and 64 (mean 45.5 years, median 46, standard deviation [SD] = ±10). Fifteen patients came for follow-up 1 year post surgery, two after 2 years, three after 3 years, two after 4 years, and two after 5 years. Thus, the time period that had passed between post surgery and follow-up ranged from 1.0 to 5.10 years. Some of the patients had undergone pitch-raising surgery on several occasions because of dehiscent suture or unsatisfactory results. All patients' voices had been recorded pre- and post surgery and at follow-up. All patients had undergone VT before surgery, as well as after surgery. Indications for surgery differed among the patients; 13 patients had problems maintaining the female pitch they had acquired in VT (5 had CTA and 8 had GP), whereas 11 patients had not been able to acquire a speaking pitch in a cisfemale voice range (see Table 2 for reference data[22]) despite VT (6 had CTA and 5 had GP). The decisions about preferred surgical techniques changed over time during the study period. In the beginning, CTA was common because it was a documented method, and GP was introduced later at our clinic. The indications for choice of surgery techniques were thus time related rather than related to indications for surgery.

Eleven of the transgender women had been operated on with CTA. Thirteen had been operated on with GP, six of them once, two patients two times, and three patients three times because the sutures had let go. Three of the patients treated with CTA had had additional GP because of inadequate results, one, two, and three times respectively. Here we report the results at follow-up after the last surgery. For those who had had both CTA and GP, the major voice change happened

after the CTA, clearly visible in the VRPs. Therefore, patients who had had CTA with additional GPs were included in the CTA group. Thus, all patients were sorted into just one of the two surgery groups (CTA and GP).

Voice therapy

All patients had VT before and after surgery. Eight patients had been treated in other parts of the country, and no information about the number of sessions or the content of the therapy program was available. For the other 16 patients, the number of sessions ranged from 5 to 20 before surgery (mean 11) and from 2 to 20 after surgery (mean 8), with weekly 45-minute sessions. VT before surgery mainly focused on raising the speaking pitch.[23] Pitch modification followed two approaches, either by trying to reach a female pitch directly or by a step-by-step increase of the pitch. Oral resonant therapy was used in attempting to change resonance through lip spreading and forward tongue placement.[24] In decreasing symptoms such as vocal fatigue and strain,[25] the accent method and semioccluded vocal tract exercises were used. Education on risk factors and vocal hygiene was brought into therapy at the very beginning. Carry-over activities were implemented at different stages. Therapy was highly individualized based on the patient's ability to adapt the different targeted parameters. After surgery, VT was focused mainly on easy and relaxed phonation, vocal hygiene education, keeping the speaking pitch in a female range, and varying the pitch to prevent a limited voice range.[26]

Surgery

Two types of surgeries had been performed during the study period: GP and CTA (see Figure 1). Both surgeries were done under general anesthesia. The GP was performed similarly to previous reports by Geneid et al[15] and Gross[27] (see Figure 1A–C). The operation was done using microlaryngoscopy. The membranous part of the VFs was measured from a scale placed on the superficial VF surface near the VF free edge. The anterior 35%–40% distance from the anterior commissure was marked with a small incision, and the superficial mucosa at the VF free edge and underside was removed with cold laryngeal microinstruments. The anterior part of the VFs was approximated with three sutures (usually PDS 5-0 or 6-0; Ethicon US LLC, Somerville, NJ). During CTA surgery, the thyroid and cricoid

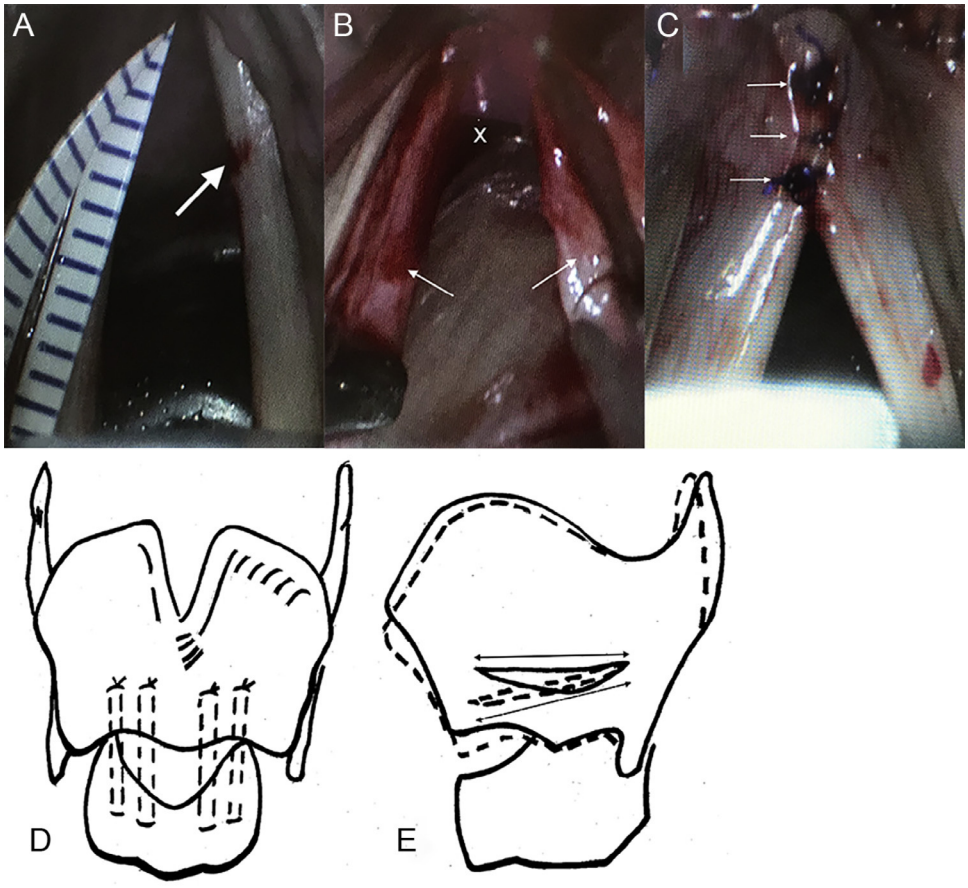

FIGURE 1. A.–E. Pitch-raising laryngeal surgery. Top: Intraoperative images of glottoplasty (vocal fold shortening): (**A**) measure scale on the left VF and on the right side cut marking the level of shortening (pexi) of the membranous VF part (*arrow*); (**B**) after removal of mucosa anterior of *arrows*, including the commissure (*marked X*); and (**C**) anterior VF membranous part approximated with three sutures (*arrows*). Bottom: Sketch of CTA procedure: (**D**) frontal view after approximation of the thyroid and the cricoid with four sutures and (**E**) lateral view before (*filled lines*) and after (*dashed lines*) approximation. The *top arrow* marks the VF length before approximation and the *lower arrow* (*dashed lines*) marks VF elongation after approximation.

cartilages were freed after a neck incision. Two holes were drilled on each side of the midline through the cricoid cartilage and 4−5 mm above the lower edge of the thyroid cartilage. The cartilages were then maximally approximated at the cricothyroid space and the sutures were tied (see Figure 1D and E). A total voice rest of 1 week was recommended after both GP and CTA. After CTA, relative voice rest was recommended for 1−2 more weeks, including avoidance of repeated coughing, throat clearing, and high physical stress. The patients were also advised initially to swallow with the chin down. No routine antibiotics were prescribed after the surgery.

Voice recordings

Digital audio recordings were made in a sound-treated recording studio at Karolinska University Hospital. The patients wore a headset with an electret microphone (Sennheiser MKE-2; Sennheiser, Wennebostel, Germany) placed at a distance of 15 cm from the mouth. The computer programs *Soundswell* and *Phog* (Neovius Data och Signalsystem AB, Lidingö, Sweden) were used for the audio recordings and the acoustic analyses. SPL was calibrated according to the program manual. The SPL values were corrected at a distance of 15 cm from mouth to microphone and were presented as if recorded at a distance of 30 cm. The threshold value for registration of phonation was set to 25 ms, also according to the program manual and the clinical routine.

A SRP was recorded (Figure 2A) while the patients read a Swedish standard text of approximately 40 seconds' duration and narrated a story from a comic strip.

Recordings of VRPs were performed to obtain information about the physiological voice range.[28] A VRP displays the range of the voice for frequency (x-axis) and voice level (y-axis). A third metric, the blackness in the picture, can show various aspects of the voice acoustics. In the present study, the third metric was the accumulated time when the patient phonated at the same frequency and decibel (Figure 2B). To produce a valid and reliable VRP, it is necessary to do the recording in a structured and careful way, which takes approximately 20−30 minutes. The guidelines described in Hallin et al[29] were followed. When recording the VRP, the patient stood in an upright position and was able to watch the VRP expand on a computer screen placed 1.5 m in front of the patient. Instructions and guidance on how to perform the VRP were given by a trained test leader throughout the recording to help the patient create the largest possible area. The patient performed the VRP on sustained phonations or glissandi solely on the vowel /a: /, commencing to phonate in the softest voice to register the lower contour of the VRP. The patient started in the midrange, decreased the pitch to register the lowest f_o first, and then increased the pitch to reach the highest f_o. The same order was used for the upper contour in the VRP. Thereafter, the patient tried to combine the lower and the upper contours to obtain a "connected" area, then to decrease the lower contour and increase the higher contour even more. Finally, the patient was asked to phonate freely to try to expand the area even further.[29]

Acoustic analyses

The following variables were extracted from the SRPs using the analysis tools in *Phog*: mean f_o (Mf_o in Hz), f_o-mode (in Hz), and range (in semitones [STs]), as shown in Figure 2A.

The following variables were extracted from the VRPs: minimum f_o (Min f_o in Hz), maximum f_o (Max f_o in Hz), and range (in ST), as shown in Figure 2B.

The original plan was to also present SPL data from the SRPs and the VRPs. However, the calibration routine had changed in the clinic over the years when data collection was performed. When investigating the decibel data, we found them unreliable. Thus, no SPL data will be presented.

Questionnaires

One questionnaire was used regularly at the hospital at that time for the study. Data were collected at the first visit (baseline), after VT, post surgery, and at follow-up. The following statements were answered on a 1−7 ordinal scale (1 = strongly disagree, 7 = totally agree): (1) *When I speak on the phone I am perceived as female*, (2) *When I speak in social contexts I am perceived as female*, (3) *I experience vocal fatigue and hoarseness when I speak*, (4) *I am satisfied with my voice*, and (5) *I am worried that my voice will reveal my native sex*.

Additional five questions were used only at follow-up. Those were chosen from a more extensive post gender-confirming pitch-raising surgery questionnaire from La Trobe University, Melbourne, Australia. Four questions were answered using a 100-mm visual analog scale (0 = not at all or never, 100 = completely or always). The questions were (1) *Are you*

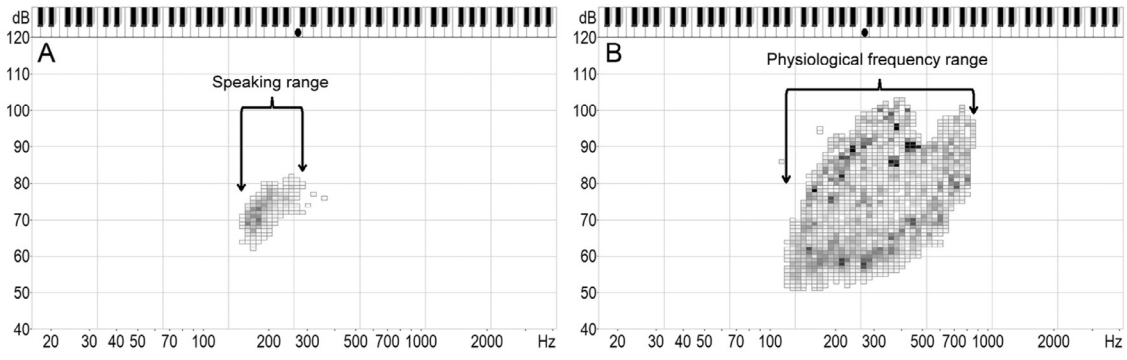

FIGURE 2. **A.−B. A.** An SRP and **B.** a VRP showing the frequencies (in Hz and semitones) on the x-axis and the sound pressure levels (in dB) on the y-axis (30-cm mouth-to microphone distance). The minimum and maximum frequencies are marked comprising the (**A**) speaking range and the (**B**) physiological frequency range.

satisfied with your voice in general? (2) *Are you satisfied with your pitch when you speak to people you know?* (3) *Are you satisfied with your pitch when you speak to people you do not know?* and (4) *Is your voice rough or hoarse in general?* The patients also answered the question: *Would you do the surgery again if you knew the surgical outcomes?* The response categories were yes and no. Neither of the questionnaires was validated.

Statistics

A mixed-effects model with repeated measures and a compound symmetry test was used to examine the effects of surgery (CTA and GP) and the effects of treatment stages over time. Acoustic data from the VRPs were from two treatment stages (before surgery and follow-up) and data from the SRPs from four treatment stages (baseline, after VT, post surgery, and follow-up). The results were based on data at baseline from 16 patients (7 in the CTA group and 9 in the GP group) and from all 24 patients at follow-up. The difference in number was due to the fact that some of the VRP recordings at baseline had not been carried out properly. The focus at that time had been on recording the Min f_o and not the Max f_o. A reliable VRP is highly dependent on the instructions to the participants, and the time spent on recording the VRPs increased over the years. It was mainly the oldest recordings, done before the clinic had implemented the procedure according to Hallin et al,[29] which needed to be excluded from the analyses. The mixed-effects model was chosen because of these missing observations. Normal distribution was assumed and confirmed with a quantile plot of residuals. The patient-reported data were rated on ordinal scales, and Friedman and Mann-Whitney tests were used to examine effects of surgery (CTA and GP) and the effects of treatment over time (baseline, after VT, and at follow-up). Assumption of equal variance of residuals across time points and surgery groups was checked and confirmed. The results in Figures 3–5 are presented using box plots showing median, quartiles (25–75%), minimum and maximum values, and outliers (>1.5 quartile range) according to the Tukey standard.

RESULTS

Acoustic f_o data—voice range profiles

The results for Min f_o, which is the lowest possible frequency a person can phonate, are based on complete data from the 24 patients and are shown in Table 1 and in Figure 3A. A significant difference between the surgery groups was found ($F(1,1) = 6.760$, $P = 0.016$), as well as a significant general effect between the treatment stages (before surgery and follow-up) ($F(1,1) = 69.251$, $P < 0.001$) and an interaction effect ($F(1,1) = 39.246$, $P < 0.001$). Pairwise comparisons showed a statistically significant increase of Min f_o for the CTA group of 38 Hz (2.3 STs), with a 95% confidence interval of 30.19–46.17, $P < 0.001$, which was not found for the GP group (5.3 Hz, 1.0 ST). The Min f_o increased for the CTA group to 123 Hz, which is within a cisfemale range according to reference cisfemale values (Table 1). The results were confirmed when the statistical analyses were performed on data in STs.

For Max f_o in the VRP, a significant general decrease was found after surgery, which is demonstrated in Figure 3B based on data in hertz ($F(1,1) = 38.771$, $P < 0.001$). Also when analyzing the data in STs, we found a significant difference between the surgery groups ($F(1,1 = 5.142$, $P = 0.034$), as well as an interaction effect ($F(1,1) = 7.927$, $P = 0.014$), meaning that the decrease was larger for the CTA group. Pairwise comparisons showed that Max f_o in the CTA group decreased with 8 STs (from 677 to 475 Hz), with a 95% confidence interval of 5.477–10.536, $P < 0.001$, and in the GP group with 3.6 STs (from 765 to 652 Hz), with a 95% confidence interval of 1.343–5.825, $P = 0.004$. For both surgery groups, the Max f_o values were very low compared with reference cisfemale values of 1275 Hz (Table 1), especially for the CTA group, with an average Max f_o of 474.7 Hz.

The *physiological frequency range* (Max f_o minus Min f_o) was calculated in STs. The results showed a general significant decrease after surgery ($F(1,1) = 150.715$, $P < 0.001$), a significant difference between the surgery groups ($F(1,1) = 7.558$, $P = 0.012$), as well as an interaction effect ($F(1,1) = 43.496$, $P < 0.001$), as seen in Figure 3C. There was no difference between the groups before surgery. Pairwise comparisons showed that the range in the CTA group decreased at 14.5 STs (from 36.8 to 22.3), with a 95% confidence interval of 16.927–11.997, $P < 0.001$. For the GP group, the range decreased at 4.4 STs (from 37.4 to 33.8), with a 95% confidence interval of 6.534–2.174, $P = 0.001$. The ranges were lower for both groups compared with reference values for cismen (40.1 STs) and ciswomen (41.2 STs) (see Table 1).

TABLE 1.
Acoustic f_o for the VRPs Presented as Mean and SD for Min f_o and Max f_o in Hertz and Range in STs Measured Before (Pre) Pitch-raising Surgery and at Follow-up for the Two Surgery Groups (CTA [n = 11] and GP [n = 13]), and Reference Data from Cismales and Cisfemales

	CTA		GP		Reference Data*	
VRP	Pre Mean (SD)	Follow-up Mean (SD)	Pre Mean (SD)	Follow-up Mean (SD)	Cismales Mean (SD)	Cisfemales Mean (SD)
Min f_o (Hz)†	84.6 (11.0)	122.8 (16.9)	87.2 (12.4)	92.5 (16.7)	74.9 (11.1)	117.7 (20.9)
Max f_o (Hz)‡	676.9 (163.2)	474.7 (206.1)	765.4 (102.5)	652.3 (121.8)	773.2 (186.1)	1275.3 (269.1)
Range (ST)‡	36.3 (4.8)	22.3 (7.1)	37.4 (3.6)	33.8 (5.2)	40.1 (4.33)	41.2 (4.22)

* Sanchez et al.[22]
† n = 24 (CTA, n = 11; GP, n = 13).
‡ n = 17 (CTA, n = 7; GP, n = 9).

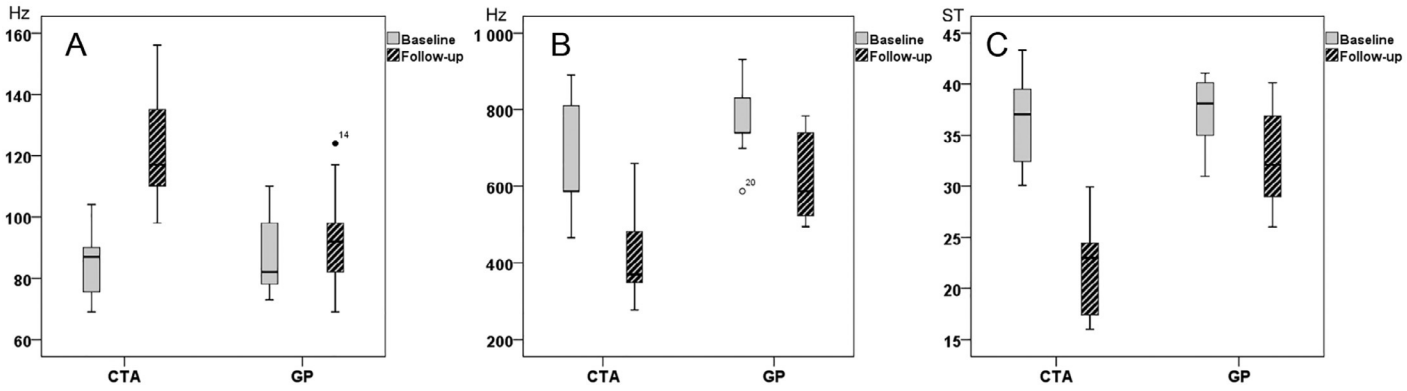

FIGURE 3. A.–C. Box plots showing results from the VRPs presented as median values and quartiles for the two surgery groups (CTA and GP) before surgery and at follow-up for (**A**) minimum f_o (Min f_o, Hz); (**B**) maximum f_o (Max f_o, Hz); and (**C**) range in semitones. The circles show outliers and the number is related to patient ID.

Acoustic f_o data—speech range profile

The Mf_o during speech increased significantly as a general treatment effect of the four measured stages (baseline, after VT, post surgery, and follow-up) ($F(1:3) = 33.091$, $P < 0.001$), and no differences were found between the surgery groups. Pairwise comparisons showed a significant increase of f_o after VT with 23 Hz (from 120.8 to 152.8), with a 95% confidence interval of −32.076 to −13.860, $P < 0.001$, as well as after surgery, with an increase of 18 Hz, with a 95% confidence interval of −26.241 to −9.746, $P < 0.001$, with no further changes at follow-up (Figure 4A). This means that f_o was stable over time after surgery. In both groups, the Mf_o had reached a gender-neutral range after VT (146.9 Hz for the CTA group and 158.6 Hz for the GP group). After surgery Mf_o increased to 170.5 and 171.9 Hz for the CTA group and the GP group, respectively, and decreased somewhat at follow-up to 167.3 and 169.4 Hz, respectively. Thus, at the follow-up, the Mf_o was within the lower frequency range for cisfemale voices according to reference values (Table 2).

The results for the f_o-mode (see Table 2) followed the same patterns as for the Mf_o, that is, a significant increase after VT, further increase after surgery, then stability over time, and no difference between the surgery groups.

The Min f_o increased significantly as a general effect of treatment stages ($F(1,3) = 57.746$, $P < 0.001$) with an interaction effect ($F(1,3) = 57.746$, $P = 0.007$), showing that the increase was larger in the CTA group (see Figure 4B). Pairwise comparisons showed that the increase happened for the CTA group after surgery with 31.3 Hz from 104.5 to 138.8 Hz, with a 95% confidence interval of −42.802 to −19.925, $P < 0.001$, and was stable after that. The Min f_o at follow-up was 134.3 Hz for the CTA group and 119.7 Hz for the GP group, both values within 1 SD from the mean reference female value of 137.1 Hz (Table 2).

For Max f_o, there was a significant difference between the surgery groups ($F(1,3) = 22.595$, $P = 0.032$), showing slightly higher values for the GP group, as seen in Figure 4C. There was also a general significant increase with treatment stages ($F(1,3) = 14.341$, $P < 0.001$). Pairwise comparisons showed that the increase happened only after VT with on average 41.8 Hz, with a 95% confidence interval of 25.782–57.899, $P < 0.001$, and was stable after that. The Max f_o values at follow-up were 232.4 Hz for the CTA group and 258.5 Hz for the GP group, which can be compared with the reference cisfemale value of 318.4 Hz (Table 2).

The *speaking range* (Max f_o minus Min f_o in the SRP), calculated in STs, differed between the surgery groups ($F(1,3) = 10.772$, $P = 0.003$) and showed a general effect of treatment stages ($F(1,3) = 13.742$, $P < 0.001$), as well as an interaction effect ($F(1,3) = 4.192$, $P = 0.022$), as seen in Figure 4D. Pairwise comparisons showed that the range was significantly reduced after surgery for the CTA group with 4.1 STs, with a 95% confidence interval of 2.455–5.841, $P < 0.001$, with no further change at follow-up. The speaking range in the CTA group decreased from 11.5 STs at baseline (via 13.5 STs after VT) to 9.5 STs at follow-up, a value considerably below the reference values for both cisfemale and cismale speakers (Table 2). The GP group increased the speaking range somewhat from 13.1 (via 14.7 STs after VT) to 13.4 STs at follow-up, which is within a cisfemale speaker's range (Table 2).

Patient-reported data

The patients rated five items, formulated as statements, on three occasions; however, not all patients responded each time because of the fact that the clinical routines were not perfectly structured in the beginning of the data collection. At the first visit, at baseline, 15 of the 24 patients responded (6 in the CTA group and 9 in the GP group), after VT 14 of the 24 patients (5 in the CTA group and 9 in the GP group), and at follow-up all 24 responded. Twelve patients answered at all three time points (5 in the CTA group and 7 in the GP group). Post surgery, only three patients answered the questionnaire; thus, these ratings were not included in the statistical analysis or shown in the figures. The results are shown in Figure 5.

The result for item 1, *When I speak on the phone I am perceived as female*, showed a general increase in the ratings and thus an improvement as an effect of treatment stages, as seen in Figure 5A. A large variation in the ratings was found in both groups especially at follow-up. The statistics indicated a significant increase in the ratings for the GP group ($P = 0.044$). For item 2, *When I speak in social contexts I am perceived as*

TABLE 2.
Acoustic f_o Data From the SRPs Presented as Mean (SD) for Mf_o, f_o-Mode, Min f_o, Max f_o in Hertz, and Range in STs Measured at the First Visit (Baseline), After VT, Post Surgery, and at Follow-up for the Two Surgery Groups (CTA [n = 7] and GP [n = 10]), and Reference Data from Cismales and Cisfemales

	CTA				GP				Reference Data*	
SRP	Baseline Mean (SD)	After VT Mean (SD)	Post Surgery Mean (SD)	Follow-up Mean (SD)	Baseline Mean (SD)	After VT Mean (SD)	Post Surgery Mean (SD)	Follow-up Mean (SD)	Cismales Mean (SD)	Cisfemales Mean (SD)
Mf_o (Hz)	118.6 (11.3)	146.9 (21.0)	170.5 (35.0)	167.3 (21.0)	138.1 (23.7)	158.6 (22.8)	171.9 (20.6)	169.4 (20.3)	109.9 (13.9)	182.8 (22.3)
f_o-mode (Hz)	107.1 (7.9)	135.0 (25.6)	161.0 (34.6)	156.2 (20.5)	127.6 (20.5)	141.8 (21.7)	158.4 (21.3)	149.4 (24.6)	-	-
Min f_o (Hz)	91.3 (7.7)	104.5 (16.3)	135.8 (31.3)	134.3 (185)	102.9 (18.1)	110.3 (22.5)	120.2 (20.1)	119.7 (21.7)	84.2 (13.9)	137.1 (19.8)
Max f_o (Hz)	178.3 (28.0)	227.7 (36.4)	231.5 (44.0)	232.4 (28.9)	212.7 (42.0)	255.8 (32.7)	264.3 (38.0)	258.5 (36.9)	181.6 (29.1)	318.4 (43.0)
Range (ST)	11.5 (2.0)	13.5 (2.4)	9.3 (2.2)	9.5 (2.2)	12.5 (3.9)	14.7 (3.5)	13.6 (2.7)	13.4 (3.5)	13.3 (3.0)	14.6 (3.2)

* Sanchez et al.[22]

female, the results showed a statistically significant effect of treatment stages for the GP group ($P = 0.018$) as demonstrated in Figure 5B. For item 3, *I experience vocal fatigue and hoarseness when I speak*, the ratings varied considerably, especially at follow-up, for both the CTA and the GP groups. The data showed a tendency for higher ratings, that is, more fatigue and hoarseness, as an effect of treatment stages in the CTA group ($P = 0.076$). The ratings for item 4, *I am satisfied with my voice*, increased significantly as an effect of treatment stages and indicated an improvement for the GP group ($P = 0.015$) and a tendency for the CTA group ($P = 0.061$). The ratings for item 5, *I am worried that my voice will reveal my native sex*, showed lower ratings as a general effect of treatment stages, interpreted as a positive outcome and improvement. For the CTA group, there was a nonsignificant improvement ($P = 0.061$), and for the GP, there was a significant improvement ($P = 0.015$).

All 24 participants also answered five additional statements and questions only at the follow-up, and the answers were possible to compare between the surgery groups. There were no statistical differences between the two groups regarding the items *Are you satisfied with your voice in general?*, *Are you satisfied with your pitch when you speak to people you know?*, and *Are you satisfied with your pitch when you speak to people you do not know?* Only on the item *Is your voice rough or hoarse in general?* did the CTA group rate significantly higher ($P < 0.05$) than the GP group. Twenty-two of the 24 patients would do the surgery again if they had known the surgical outcomes. One patient did not answer. One patient who had been treated with CTA and additional GPs had ended up with an extremely limited pitch range in both the VRP and the SRP and responded definitely no.

DISCUSSION

In the present study, we have compared results from two pitch-raising surgery techniques, which are commonly used today in many ear, nose, and throat clinics around the world. Acoustic as well as patient-reported data were collected at baseline, after VT, post surgery, and at follow-up at least 1 year after the surgery was performed. Thus, we have been able to compare long-term results from two surgery methods and to describe the advantages and disadvantages of those methods, which have not been done in a systematic way before. The discussion is organized based on the research questions and methodological issues.

Research question 1: Is the physiological frequency range affected by pitch-raising surgery? If so, how much and has it reached cisfemale range?

We found it important to document what happens with the physiological voice range to understand the prerequisites for the function of speaking voice. Our results showed a significant increase of Min f_o in the VRP, that is, the lowest possible phonation the person can produce. This is a highly desirable result because many patients have problems with pitch drops to cismale frequencies (eg, see Reference 30). The increase was found only after CTA and not after GP. However, CTA also had a very negative result, namely, a significant decrease of the Max f_o in the VRP, that is, the highest possible phonation a person can produce. It dropped by 8 STs to an average of 474.7 Hz, which

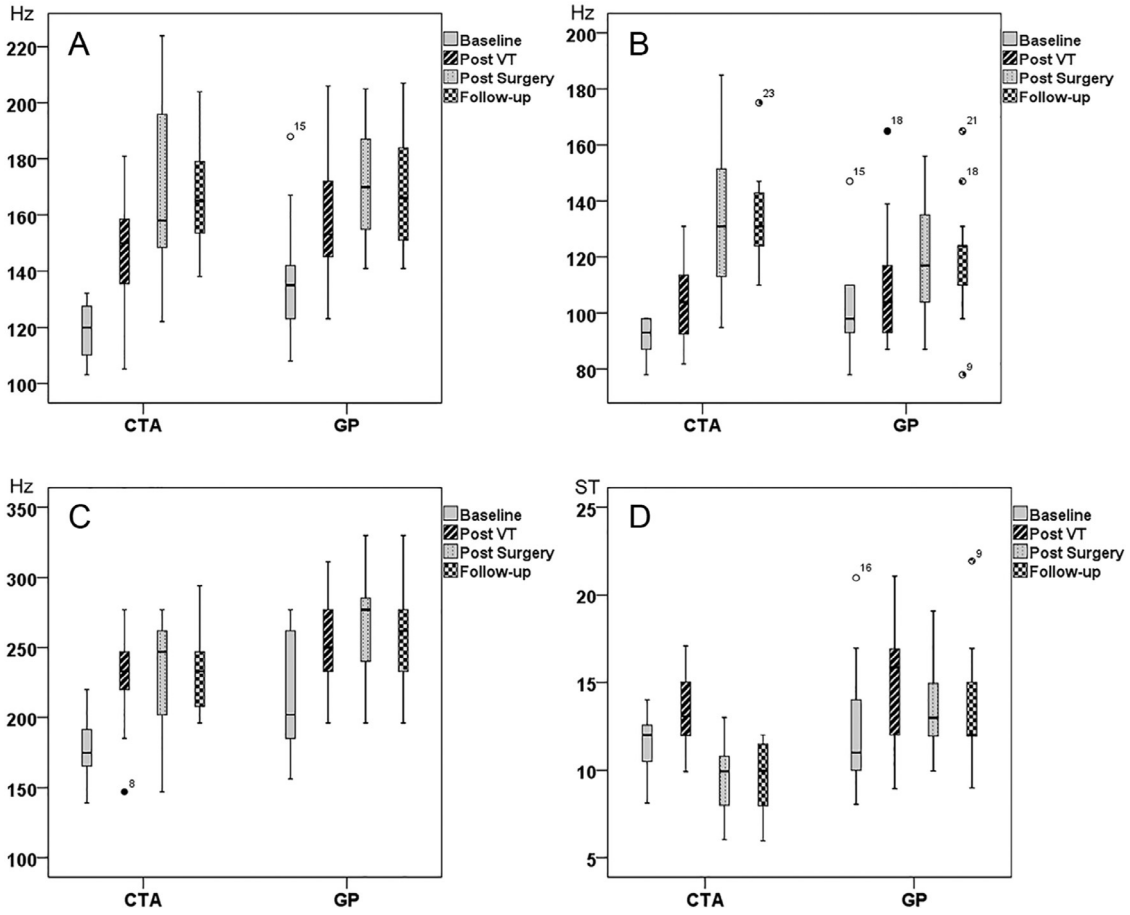

FIGURE 4. **A.–D.** Box plots showing the results from the SRPs presented as median values and quartiles for the two surgery groups (CTA and GP) at baseline, after VT, post surgery, and at follow-up for (**A**) mean f_o (Mf_o, Hz); (**B**) minimum f_o (Min f_o, Hz); (**C**) maximum f_o (Max f_o, Hz); and (**D**) speaking range in STs. The circles show outliers and the number is related to patient ID.

is very low compared with both cismale (723 Hz, SD 186) and cisfemale (1275 Hz, SD 269) reference values.[22] Pitch variation is generally important if a speaking voice is to sound natural and is especially important in allowing transgender women to sound female. When the physiological range became highly restricted, as happened to some patients (but only after CTA), this was regarded as a highly undesirable complication after surgery. No perceptual analysis was done in the present study, but the degree of dysphonia was reported as significantly higher in the CTA group than in the GP group based on the patient-reported data at follow-up. Three patients were not satisfied with the results after CTA because of dysphonia and a speaking pitch that remained too low after surgery, and therefore GP was performed as well. Complications involving restricted voice range and dysphonia after CTA have also been reported by Kocak et al,[19] who tried to treat the VFs and improve the voice with laser surgery after CTA, with satisfactory results. In the present study, the maximum pitch in the VRP was lowered for the GP group too, but not as much, 3.4 STs. The CTA technique stretches the VFs to obtain a higher pitch. This finding may explain why the Max f_o decreases because further stretching is not possible. It is more difficult to understand the lowering of Max f_o after GP, but one assumption is that the GP technique can cause increased stiffness in the VFs affecting the overall vocal dynamics, both in frequency and SPL.

Previous studies have reported data on the total frequency range before and after surgery[10,14,31]; however, those data differ from ours. One likely explanation is that the procedures and instructions to the participants differ between the studies. It is especially difficult to compare results when the procedures and methods for measuring pitch range have not been described in detail. It has been stated that when the physiological frequency range is measured using VRPs, the results are highly dependent on the instructions to and coaching of the participant, as well as on the time spent for the VRP.[22,28,29] We realized that some VRP measurements in the present study were not carried out properly, and thus we had to exclude those recordings when calculating Max f_o and the range from the VRPs. Some VRPs did not look complete when examined visually, and this impression was confirmed by the short recording time, such as 5–10 minutes, that had been logged in the data file. Recordings lasting 20–30 minutes are recommended for reliable VRP data.[29] This also includes documentation of the dynamics of the voice, that is, minimum and maximum SPL across the frequency range. Those data were not reported in the present study; not all of the recordings were considered

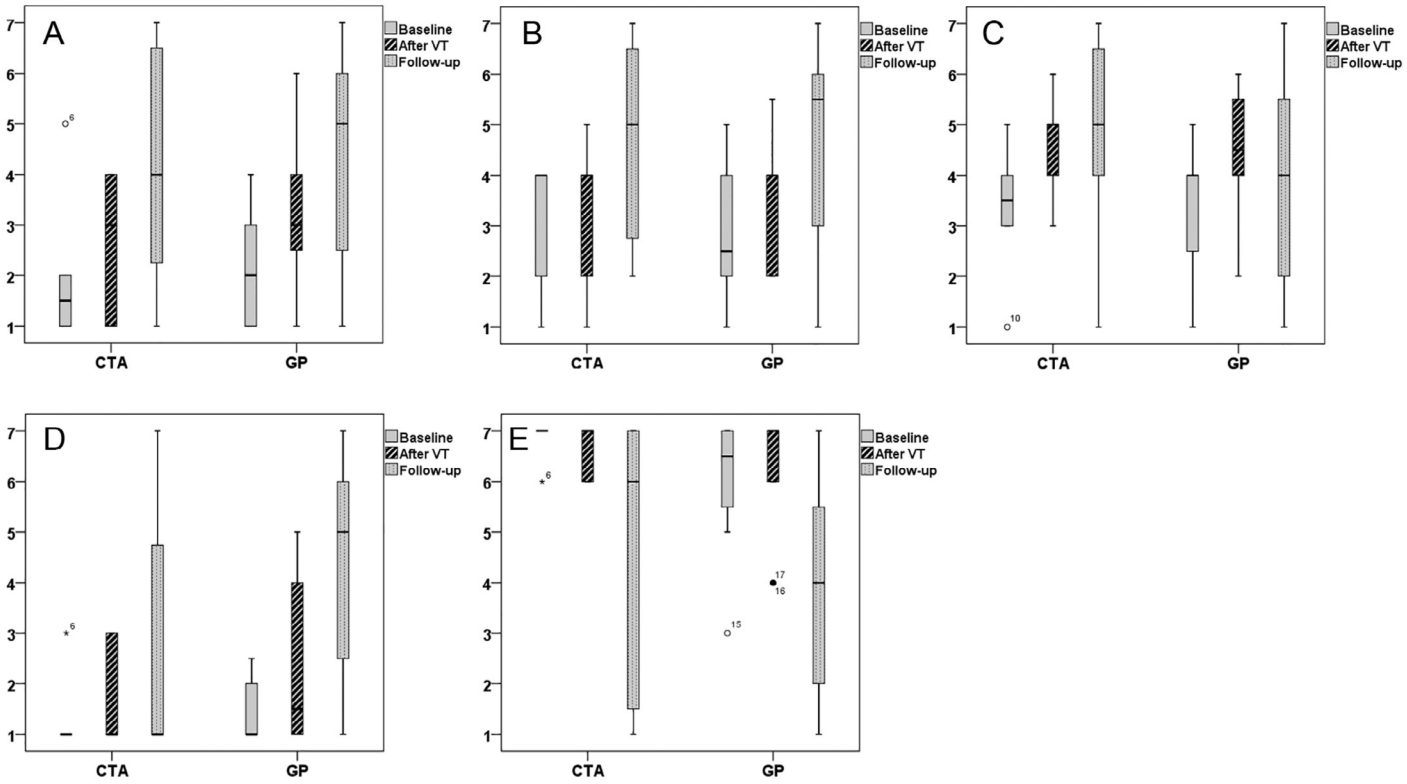

FIGURE 5. **A.–E.** Box plots showing median values and quartiles for patient-reported data for the two surgery groups (CTA and GP) at baseline, after VT, and at follow-up, for the statements (**A**) "When I speak on the phone I am perceived as female," (**B**) "When I speak in social contexts I am perceived as female," (**C**) "I experience vocal fatigue and hoarseness when I speak," (**D**) "I am satisfied with my voice," and (**E**) "I am worried that my voice will reveal my native sex." (1 = totally disagree, 7 = totally agree). The circles show outliers and the number is related to patient ID.

reliable, mainly because of changes in the calibration procedure for SPL during the years of data collection.

Research question 2: Is speaking f_o affected by pitch-raising surgery? If so, how much and has it reached cisfemale range?

The speaking f_o, measured with Mf_o and f_o-mode, was raised significantly after both CTA and GP surgery in agreement with previous studies.[8,10,14,16,18] We also found that Mf_o and f_o-mode were stable at follow-up, again in agreement with others.[10,18] When interpreting the results, it is relevant to compare the f_o values to reference cisfemale values. At follow-up, both surgery groups in our study had reached Mf_o values (CTA 167.3 Hz and GP 169.4 Hz) within the lower interval of a cisfemale frequency range (182.8 Hz, SD 22.3).[22] In many other studies, only the change in hertz or STs is presented, which makes the results difficult to interpret. In the present study, Mf_o and f_o-mode increased significantly after VT to a gender-neutral range (140–160 Hz); however, neither the patients nor the clinicians were satisfied, and that is why the surgeries were performed. A majority of our patients at Karolinska University Hospital do not receive surgery because they are satisfied after VT or are aware of the risk of complications after surgery. One important finding was that maximum f_o in the SRP increased significantly only after VT and not further after surgery. One part of VT is to practice varying the pitch upward and to avoid phonating at a low pitch to adapt a female voice.[30] The Min f_o increased, as a general effect of treatment stages in both surgery groups, indicating that it is easier for the patient to avoid pitch drops after surgery when speaking. Raising the minimum pitch of the speaking voice, sometimes called "the pitch floor," has been reported as one aspect that helps ensure that a transgender voice is being successfully perceived as female compared with less successful results when the voice is still perceived as male.[32] The speaking range became significantly more restricted in the CTA group than in the GP group, a finding caused by the decrease of the Max f_o and not by an increase of the Min f_o. This cannot be regarded as a positive finding for CTA: the result is in favor of GP. All patients had received VT after surgery. The rationale for postoperative VT is that the VFs need training to be flexible after voice rest and surgery. However, the efficacy of VT post surgery needs to be investigated in future well-designed trials.

Research question 3. Are the patients satisfied with the voice after surgery?

One question in the present study was directly posed about this. There was a significant increase in satisfaction with the voice as a general effect of treatment stages. When the surgery groups were compared, those who had undergone GP seemed more satisfied. The ratings ranged from the lowest to the highest values for both surgery groups at follow-up, indicating that some participants found surgery and VT very helpful and others did not. The findings also suggest that patient-reported outcomes are not directly associated with instrumental measures. The results from the other statements about patient satisfaction with the voice showed significant improvements with

treatment stages. Also, 22 of the 24 patients said they would do the surgery again if they had known the outcomes ahead of time. However, one patient treated with CTA ended up having a very restricted pitch range, as shown in both the VRP and the SRP data. The patient also gave high ratings on dysphonia and regretted having done the surgery. It should be remembered that the CTA-technique is irreversible, so when the results are not successful, the patient may pay a high price. The GP, on the other hand, may be repeated several times to get the desired effect. Thus, the repeated number of GP surgeries should not only be regarded as negative. During the period of the study, we found that the patients needed more VF shortening to get an adequate effect on voice. Also, the dimensions and types of sutures were changed over time. A majority of the participants in the present study were satisfied, in accordance with results from the recent surveys by Van Damme et al[8] and by Song and Jiang.[20] Thus, surgery certainly is an important option for many patients when VT alone has not been successful.

Research question 4: Are there voice-related complications after surgery?

The answer to this question is based on results from acoustic and patient-reported data. Vocal fatigue and hoarseness tended to increase after surgery in the CTA group but not in the GP group. However, it is not clear which one of the symptoms the patient actually meant. Our clinical experience is that it is mainly hoarseness that increases after surgery. Thus, this question needs to be divided in two in the future. Patients operated on with CTA rated their dysphonia as significantly higher than those operated on with GP based on ratings at follow-up. This result differs from the results of other studies, which have found dysphonia among patients after GP,[11] and studies reviewed in Van Damme et al.[8] However, as long as the person perceives that she has a female voice, dysphonia may not be so important, depending on the vocal demands. In the present study, we did not do an audio-perceptual analysis, so our results on dysphonia are based on patient-reported data. Further studies are needed to ask the patients to elaborate on their experiences and how they value voice quality vs pitch. Based on our clinical experience, many patients' voices are weaker after surgery and our aim was to also measure SPL in the SRP and VRP. However, this was not possible because of technical problems with the calibration procedure over the years. We recommend that measures of SPL be included in future follow-up studies. Better validated questionnaires are needed to shed light on the problem area. Since this data collection started in 2004, the Transsexual Voice Questionnaire has been translated to Swedish and validated and is now used routinely at our clinic. That questionnaire was developed because the Voice Handicap Index had been found not to grasp the particular voice issues that transgender women experience.[33]

Research question 5. What surgical method should be recommended?

There are advantages and disadvantages with both methods. There were no differences regarding average speaking pitch between the methods, but the speaking range was more restricted in the CTA group. The differences were significant regarding Min f_o in the VRP; Min f_o became significantly higher after CTA, which is very much desired. However, the negative effects after CTA of a very low Max f_o in the VRP and extremely restricted pitch range during speaking must be considered severe complications. CTA surgery is irreversible, and if the results are not satisfactory, the patient must live with a restricted voice for the rest of her life. Therefore, we recommend that CTA be avoided or carefully discussed before being chosen. If an individual has an extremely low Min f_o in the VRP, CTA may be considered. The GP has its advantages and disadvantages too. The surgery can be repeated several times, which means that the VFs can be shortened even more after the first surgery, if necessary. On the other hand, sutures can become dehiscent, which is a disappointment for the patient, and repeated surgery means increased costs for the health-care system and increased risk for the patient, who must undergo repeated general anesthesia. As a consequence of the results of the present study, CTA is no longer performed at Karolinska University Hospital. When comparing CTA with GP, it is clear that the GP has a potential for development, which is not the case for CTA. Therefore, we recommend further development of the GP technique, in agreement with Song and Jiang.[20] Successful results have recently been reported based on a large set of data from patients treated with a technique that combines glottoplastic shortening of the VFs with a retrodisplacement of the anterior commissure Vocal Fold with RetrodiSplacement of the Anterior Commisure.[16] Therefore, the technique that is under implementation at our hospital is the Vocal Fold with RetrodiSplacement of the Anterior Commisure,[16] in combination with shortening the VFs by 40%−50%.

In the future, it is important and necessary to formalize the content of the VT treatment programs both pre- and post surgery, for example, how to achieve target goals, so that research studies can be carried out in a more systematic way than what has been done before. We need prospective treatment studies for solid outcomes when studying the efficacy of VT, as well as pitch-raising surgery. Furthermore, there is a strong need for criteria for decision-making if and when pitch-raising surgery should be carried out after VT. With those criteria as a basis and improved surgery techniques, we will be able to be more confident to decide at what stage patients should be offered pitch-raising surgery.

CONCLUSIONS

There are advantages and disadvantages with both CTA and GP. The results of the present study can be used as a basis for surgical decision-making about individual patients. Only a minority of the transgender women at our hospital require treatment with surgery after VT. But those who need surgery are more satisfied with their voices after surgery, despite dysphonia. The CTA increases Min f_o in the physiological frequency range, which is desirable, but reduces Max f_o and the physiological and speaking range, which do not favor an optimal female vocal behavior. Thus, CTA is no longer performed at our hospital, whereas surgical techniques based on GP are

being developed further. We strongly recommend the use of VRPs to evaluate treatment effects.

Acknowledgments

The authors acknowledge the financial support from Aina Börjeson's Fund and from Karolinska University Hospital, which made it possible for the main author to analyze data and to prepare the manuscript. The authors thank Janet Holmén for language revision, Svante Granqvist for help with acoustic analysis and preparing the figures, and Magnus Backheden for statistical support.

REFERENCES

1. Dhejne C, Lichtenstein P, Boman M, et al. Long-term follow-up of transsexual persons undergoing sex reassignment surgery: cohort study in Sweden. *PLoS ONE*. 2011;6 e16885.
2. Dhejne C, Öberg K, Arver S, et al. An analysis of all applications for sex reassignment surgery in Sweden, 1960−2010: prevalence, incidence, and regrets. *arch sex behav*. 2014;43:1535–1545.
3. Andrews ML, Schmidt CP. Gender presentation: perceptual and acoustical analyses of voice. *J Voice*. 1997;11:307–313.
4. Dacakis G, Oates J, Douglas J. Beyond voice: perceptions of gender in male-to-female transsexuals. *Curr Opin Otolaryngol Head Neck Surg*. 2012;20:165–170.
5. Van Borsel J, Janssens J, De Bodt M. Breathiness as a feminine voice characteristic: a perceptual approach. *J Voice*. 2009;23:291–294.
6. Oates J. Evidence-based practice in voice therapy for transgender/transsexual clients. In: Adler RK, Hirsch S, Mordaunt M, eds. *Voice and Communication Therapy for the Transgender/Transsexual Client: a Comprehensive Clinical Guide*. San Diego, CA: Plural Pub; 2012:45–68.
7. Oates J, Dacakis G. Transgender voice and communication: research evidence underpinning voice intervention for male-to-female transsexual women. *SIG 3 Perspect Voice Voice Dis*. 2015;25:48–58.
8. Van Damme S, Cosyns M, Deman S, et al. The effectiveness of pitch-raising surgery in male-to-female transsexuals: a systematic review. *J Voice*. 2017;31:244 e241-244.e245.
9. Isshiki N, Taira T, Tanabe M. Surgical alteration of the vocal pitch. *J Otolaryngol*. 1983;12:335–340.
10. Neumann K, Welzel C. The importance of the voice in male-to-female transsexualism. *J Voice*. 2004;18:153–167.
11. Mastronikolis NS, Remacle M, Biagini M, et al. Wendler glottoplasty: an effective pitch raising surgery in male-to-female transsexuals. *J Voice*. 2013;27:516–522.
12. Wendler J. *Vocal Pitch Elevation After Transsexualism Male to Female*. Salsomaggiore, Italy. 1990 Proceedings of the Union of the European Phoniatricians.
13. Orloff LA, Mann AP, Damrose JF, et al. Laser-assisted voice adjustment (LAVA) in transsexuals. *Laryngoscope*. 2006;116:655–660.
14. Thomas JP, Macmillan C. Feminization laryngoplasty: assessment of surgical pitch elevation. *Eur Arch Otorhinolaryngol*. 2013;270:2695–2700.
15. Geneid A, Rihkanen H, Kinnari TJ. Long-term outcome of endoscopic shortening and stiffening of the vocal folds to raise the pitch. *Eur Arch Otorhinolaryngol*. 2015;272:3751–3756.
16. Kim HT. A new conceptual approach for voice feminization: 12 years of experience. *Laryngoscope*. 2017;127:1102–1108.
17. Brown M, Perry A, Cheesman AD, et al. Pitch change in male-to-female transsexuals: has phonosurgery a role to play? *Int J Lang Commun Disord*. 2000;35:129–136.
18. Kanagalingam J, Georgalas C, Wood GR, et al. Cricothyroid approximation and subluxation in 21 male-to-female transsexuals. *Laryngoscope*. 2005;115:611–618.
19. Kocak I, Akpinar ME, Cakir ZA, et al. Laser reduction glottoplasty for managing androphonia after failed cricothyroid approximation surgery. *J Voice*. 2010;24:758–764.
20. Song TE, Jiang N. Transgender phonosurgery: a systematic review and meta-analysis. *Otolaryngol Head Neck Surg*. 2017;156:803–808.
21. Södersten M, Nygren U, Hertegård S, et al. Interdisciplinary program in Sweden related to transgender voice. *SIG 3 Perspect Voice Voice Disord*. 2015;25:C1–C4.
22. Sanchez K, Oates J, Dacakis G, et al. Speech and voice range profiles of adults with untrained normal voices: methodological implications. *Logoped Phoniatr Vocol*. 2014;39:62–71.
23. Dacakis G. Long-term maintenance of fundamental frequency increases in male-to-female transsexuals. *J Voice*. 2000;14:549–556.
24. Carew L, Dacakis G, Oates J. The effectiveness of oral resonance therapy on the perception of femininity of voice in male-to-female transsexuals. *J Voice*. 2007;21:591–603.
25. Söderpalm E, Larsson A, Almquist SA. Evaluation of a consecutive group of transsexual individuals referred for vocal intervention in the west of Sweden. *Logoped Phoniatr Vocol*. 2004;29:18–30.
26. Mount KH, Salmon SJ. Changing the vocal characteristics of a postoperative transsexual patient: a longitudinal study. *J Commun Disord*. 1988;21:229–238.
27. Gross M. Pitch-raising surgery in male-to-female transsexuals. *J Voice*. 1999;13:246–250.
28. Ternström S, Pabon P, Södersten M. The voice range profile: its function, applications, pitfalls and potential. *Acta Acustica United Acustica*. 2016;102:268–283.
29. Hallin AE, Fröst K, Holmberg EB, et al. Voice and speech range profiles and Voice Handicap Index for males—methodological issues and data. *Logoped Phoniatr Vocol*. 2012;37:47–61.
30. Wolfe VI, Ratusnik DL, Smith FH, et al. Intonation and fundamental frequency in male-to-female transsexuals. *J Speech Hear Disord*. 1990;55:43–50.
31. Anderson JA. Pitch elevation in transgender patients: anterior glottic web formation assisted by temporary injection augmentation. *J Voice*. 2014;28:816–821.
32. Södersten M, Holmberg EB. *Speech range profiles for male-to-female transsexual patients before and after voice therapy*. Marseille, France. 2011 Oral presentation at the 9th Pan European Voice Conference (PEVOC), September 2nd.
33. Dacakis G, Davies S, Oates JM, et al. Development and preliminary evaluation of the transsexual voice questionnaire for male-to-female transsexuals. *J Voice*. 2013;27:312–320.