Facial Recognition Neural Networks Confirm Success of Facial Feminization Surgery

Kevin Chen, M.D.
Stephen M. Lu, M.D., M.Div.
Roger Cheng, M.S.
Mark Fisher, M.D.
Ben H. Zhang, B.A.
Marcelo Di Maggio, M.D.
James P. Bradley, M.D.

New York, N.Y.; and Buenos Aires, Argentina

Background: Male-to-female transgender patients desire to be identified, and treated, as female, in public and social settings. Facial feminization surgery entails a combination of highly visible changes in facial features. To study the effectiveness of facial feminization surgery, we investigated preoperative/postoperative gender-typing using facial recognition neural networks.

Methods: In this study, standardized frontal and lateral view preoperative and postoperative images of 20 male-to-female patients who completed hard- and soft-tissue facial feminization surgery procedures were used, along with control images of unoperated cisgender men and women ($n = 120$ images). Four public neural networks trained to identify gender based on facial features analyzed the images. Correct gender-typing, improvement in gender-typing (preoperatively to postoperatively), and confidence in femininity were analyzed.

Results: Cisgender male and female control frontal images were correctly identified 100 percent and 98 percent of the time, respectively. Preoperative facial feminization surgery images were misgendered 47 percent of the time (recognized as male) and only correctly identified as female 53 percent of the time. Postoperative facial feminization surgery images were gendered correctly 98 percent of the time; this was an improvement of 45 percent. Confidence in femininity also improved from a mean score of 0.27 before facial feminization surgery to 0.87 after facial feminization surgery.

Conclusions: In the first study of its kind, facial recognition neural networks showed improved gender-typing of transgender women from preoperative facial feminization surgery to postoperative facial feminization surgery. This demonstrated the effectiveness of facial feminization surgery by artificial intelligence methods. (*Plast. Reconstr. Surg.* 145: 203, 2020.)

CLINICAL QUESTION/LEVEL OF EVIDENCE: Therapeutic, IV.

Facial feminization surgery is often an important aspect of gender confirmation. In the late 1980s, Ousterhout used an understanding of craniofacial anatomy and realization of the specific differences between the male and female facial skeletons to optimize facial feminization surgery procedures.[1] For many patients, feminizing the face is an even more important step to their journey of reaching their desired gender identity than "top" (breast augmentation) or "bottom" (vaginoplasty) operations. Being identified as female in everyday exchanges with the public is of utmost importance. In these daily interactions, the face is the main visible feature determining gender and, frequently, despite years of hormonal therapy and expert application of makeup, hair, or wigs, patients are often still misidentified as male.

Coincidentally, not long after the first feminization procedures were being performed, researchers applied computers to the task of gender

From the Division of Plastic and Reconstructive Surgery, Zucker School of Medicine at Hofstra/Northwell; Microsoft Corp.; and the Division of Plastic and Reconstructive Surgery, British Hospital of Buenos Aires.
Received for publication January 27, 2019; accepted May 16, 2019.
The authors listed first and second contributed equally and are co–first authors.

DOI: 10.1097/PRS.0000000000006342

Disclosure: *The authors have no financial disclosures.*

recognition through the use of convolutional neural networks, a series of self-learning algorithms that are used to identify images. Successive layers of the neural network recognize patterns of decreasing granularity, beginning at the pixel level, and progressing through patterns/edges/lines to shapes, and ultimately to faces (Fig. 1). These networks are fed millions of training images to develop and refine the models, with backpropagating algorithms that iteratively optimize the parameters to produce the most accurate model. The network is not told what or how to identify specific features; it autonomously determines what image patterns are most likely to match the aggregate information derived from the millions of training images. With the advantage of reviewing millions of faces, a neural network has potentially "more experience" than the typical human who sees tens of thousands of faces over the course of a lifetime.

Neural networks were being used for identification of faces as male or female as early as 1991 by Gray et al.[2] Even in its infancy as a tool for determining gender, neural networks were able to outperform humans in gender discrimination. With the rapid increase in computing power at a decreased cost over the past decade, modern deep convolutional neural networks are even more powerful at face detection and gender classification and have commercial applications ranging from facial identification to unlock phones to natural language processing and speech recognition.[3] Very recently, DeepGestalt (FDNA, Inc., Boston, Mass.), a facial recognition software, was shown to be superior in the recognition of known genetic syndromes based on facial features to the capabilities of clinical experts.[4]

We wanted to use modern neural networks to confirm successful facial feminization surgery. To do this, we used four of the most sophisticated neural network programs to analyze preoperative and postoperative facial feminization patients and identify gender. Male and female controls were also used. To our knowledge, there has been no prior research on facial feminization outcomes through the use of machine learning and neural networks.

PATIENTS AND METHODS

Frontal and lateral view images of the face were standardized and processed for neural network determination of gender ($n = 120$). Preoperative and postoperative images of 20 consecutive patients from 2013 to 2018 who had completed all stages of facial feminization surgery performed by either of the two senior surgeons (J.P.B. and M.D.M.) and consented to the study were used. For controls, images of 10 cisgender unoperated male and 10 cisgender unoperated female patients were used. All facial feminization surgery was performed following approval of a multidisciplinary team and successful completion of psychosocial evaluations. Patients underwent staged facial feminization surgery procedures. Hard-tissue feminization typically included frontal sinus setback (all study patients), bilateral lateral supraorbital rim reduction by ostectomy and anteroposterior burring (95 percent), bilateral mandibular angle reduction by resection and burring (85 percent), osseous genioplasty with narrowing (80 percent) and/or vertical height reduction (65 percent), and laryngoplasty/tracheal shave (45 percent).

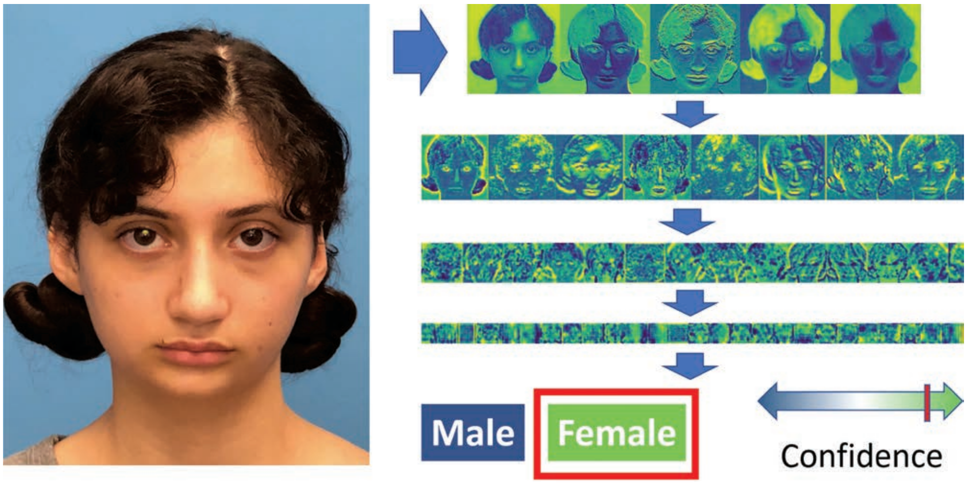

Fig. 1. Illustration of neural network processing from the facial image input on the *left* to the pixilated isolation of shapes, lines, angles, colors, to final output of gender identification and confidence reporting (*below*, *right*).

Soft-tissue feminization typically included brow lift/shortening (80 percent), septorhinoplasty (75 percent), upper lip shortening (55 percent), and fat grafting (80 percent). Less commonly used procedures included orthognathic surgery (double jaw) (10 percent), malar augmentation (10 percent), face lifting (10 percent), and blepharoplasty (5 percent).

All patient and control images were each processed through four convolutional neural networks: (1) Microsoft, (2) IBM, (3) Amazon, and (4) Face++, and the gender was identified. The Amazon and IBM software also generated a confidence level from 0 to 1 associated with their gender classification (with 0 being the least confident and 1 being the most confident). Three points of analysis were important to the study: (1) correct identification of gender versus misidentification for each of the four groups (i.e., male controls, preoperative facial feminization surgery, postoperative facial feminization surgery, and female controls); (2) improvement of gender identification from preoperative facial feminization surgery to postoperative facial feminization surgery; and (3) confidence of femininity for each of the four groups.

Statistical analysis was then performed with Stata 14 (StataCorp, College Station, Texas). Chi-square analysis was used to compare the proportion of preoperative and postoperative photographs that were classified as female. To scale and compare confidence for photographs that were preoperatively identified as male and postoperatively identified as female, photographs that were identified as male were assigned a negative confidence. The scale then ranged from −1 to 1, with −1 being very confident that the patient was male and 1 being very confident that the patient was female.

RESULTS

Demographics of the four groups were similar with regard to age and race (Table 1). A total of 480 responses were requested for 120 photographs from four neural networks capable of facial recognition and gender identification (i.e., Microsoft, IBM, Amazon, and Face++). Microsoft's network was not trained to recognize facial lateral views, and there were nine other instances where a face was not recognized by the neural network, giving a total of 411 responses.

First, neural networks correctly identified the gender of cis-male controls 98.6 percent of the time (Table 2). Cis-female controls were correctly identified 91.2 percent of the time and misidentified only 8.8 percent of the time. However, preoperative facial feminization surgery images were misidentified 41.4 percent of the time (recognized as male) and correctly identified as female only 58.6 percent of the time. By contrast, postoperative facial feminization surgery images were identified correctly 93.7 percent of the time; this was similar to cis-female controls. Likewise, neural network recognition of frontal images as female was as follows: male controls, 0 percent; preoperative facial feminization surgery, 53 percent; postoperative facial feminization surgery, 98 percent; cis-female controls, 98 percent (Fig. 2). There were differences in gender recognition in frontal versus lateral images (mean difference range, 3 to 17 percent) (Table 2). Neural networks are better trained in frontal facial as opposed to lateral facial images.

Second, for every individual facial feminization surgery study patient there was an improvement in gender recognition by the neural networks. There was a mean improvement of 45 percent in correct gender recognition from preoperative facial feminization surgery frontal images to postoperative facial feminization surgery frontal images (Fig. 3). Postoperative facial feminization surgery images were more likely to be classified as female compared to preoperative facial feminization surgery images across all software packages and all views (both frontal and lateral) (Table 2).

Third, for confidence levels of the gender selection, male control images showed a confidence of -0.912 ± 0.07 and cis-female control

Table 1. Neural Network Study Patient Demographics

		Race						
Group	Age (yr)	White (%)	African American (%)	Hispanic (%)	Asian (%)	American Indian (%)	Pacific Islander (%)	Other (%)
Male controls	35	50	20	20	10	0	0	0
Preoperative FFS	37	60	20	10	5	5	0	0
Postoperative FFS	37	60	20	10	5	5	0	0
Female controls	37	60	20	10	10	0	0	0

FFS, facial feminization surgery.
*Demographics of study patients showed similar age and race among the four groups.

Table 2. Percentage of Patient Images Identified as Female by Each Neural Network

Neural Network and Image View	Male Control (%)	Preoperative FFS (%)	Postoperative FFS (%)	Female Control (%)	*p* (Preoperative vs. Postoperative)
Amazon					
Frontal	0	60	95	90	0.008*
Lateral	10	75	100	100	0.017*
IBM					
Frontal	0	60	100	100	0.002*
Lateral	0	65	95	100	0.018*
Face++					
Frontal	0	40	95	100	<0.001*
Lateral	0	30	70	44	0.011*
Microsoft					
Frontal	0	50	100	100	<0.001*
Lateral	†	†	†	†	†
Mean					
Frontal	0	53	98	98	<0.001*
Lateral	3	68	89	81	<0.001*

FFS, facial feminization surgery.
*$p < 0.5$.
†The Microsoft neural network was not trained to recognize faces from a lateral view.

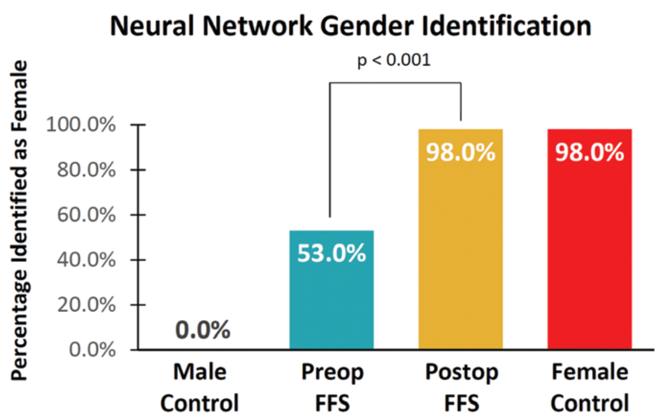

Fig. 2. Bar graph depicts percentage of frontal images identified as female. All neural networks consistently identified female controls as female and male controls as not female (i.e., identified as male). Preoperative facial feminization surgery (*FFS*) patients were identified as female only approximately half the time, but postoperative facial feminization surgery patients were identified as female at a rate comparable to female controls. The difference between preoperative facial feminization surgery and postoperative facial feminization surgery was significant ($p < 0.01$).

showed an average confidence of 0.887 ± 0.09 (Fig. 4). Both the IBM- and Amazon-generated confidence levels showed significant improvement from preoperative facial feminization surgery to postoperative facial feminization surgery images for both frontal and lateral views (Table 3). The greatest changes in confidence were on Amazon analysis of lateral view photographs, which demonstrated an average preoperative confidence of 0.19 ± 0.35 and postoperative confidence of 0.83 ± 0.06 ($p < 0.001$), and on IBM analysis of frontal view photographs, which demonstrated an average preoperative confidence of 0.25 ± 0.43 and average postoperative confidence of 0.98 ± 0.02 ($p < 0.001$).

DISCUSSION

Facial feminization surgery is deemed successful when transgender women are more likely to be identified as female based on facial appearance and aesthetics. Although certain aspects of facial appearance can be altered with hormonal therapy, makeup, and wigs, facial feminization surgery addresses fundamental anatomical differences between male and female faces that

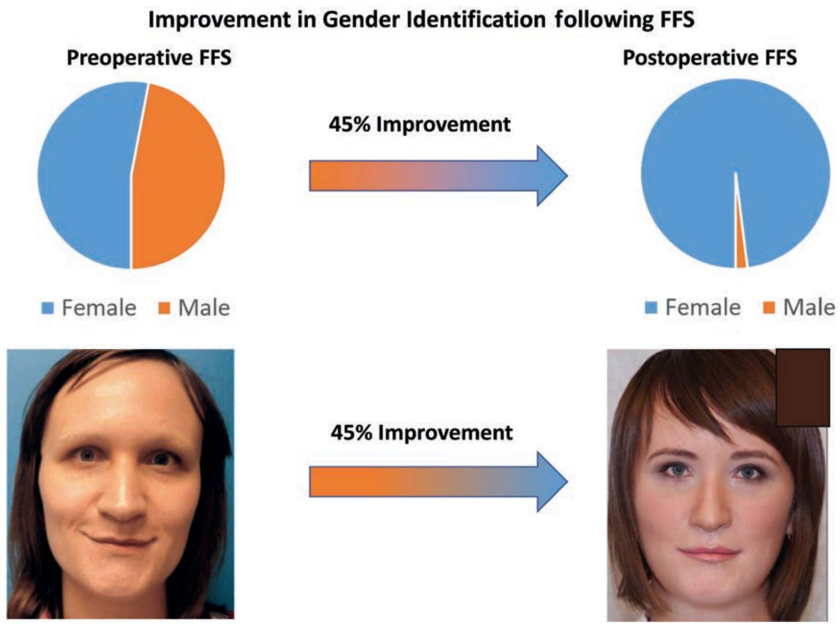

Fig. 3. Pie chart shows the improvement of gender identification from preoperative facial feminization surgery (*FFS*) to postoperative facial feminization surgery. There was significantly less misgendering after facial feminization surgery compared to before facial feminization surgery.

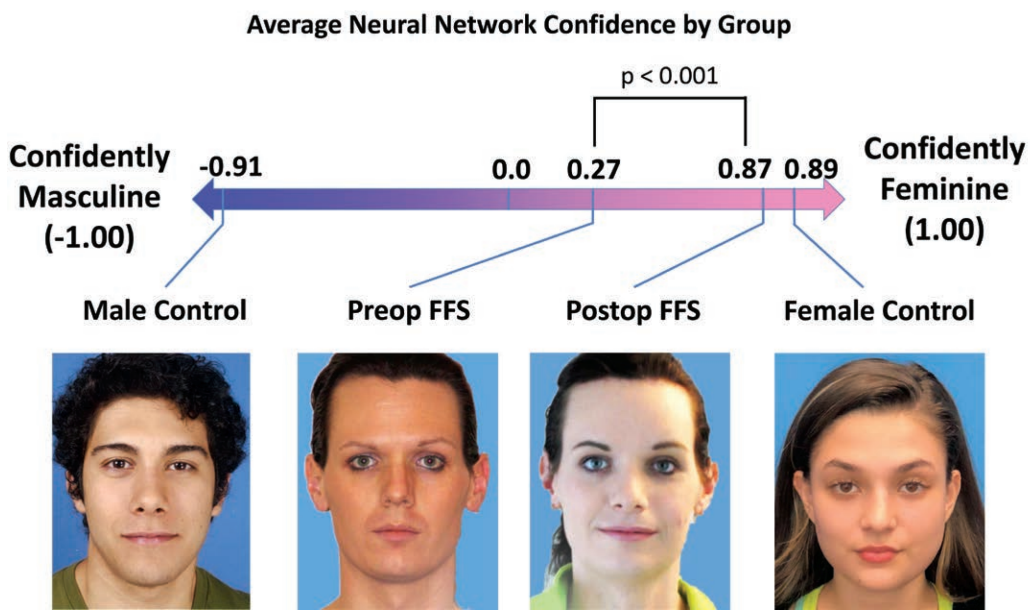

Fig. 4. Neural network spectrum of confidence in gender identification (from −1.0 confidence in masculinity to 1.0 confidence in femininity) showed progressive increased confidence in femininity from male controls to preoperative facial feminization surgery (*FFS*) to postoperative facial feminization surgery to female controls.

develop during childhood and puberty.[2,3] Raffaini et al. showed good patient satisfaction following various facial feminization surgery procedures in 33 patients.[5] In addition, Ainsworth and Spiegel showed quality-of-life improvement for patients after facial feminization surgery.[6] However, there is a recognized need for standardized patient-reported outcome measures and objective measures of efficacy.[7–9]

We sought to use the recent advances in modern deep neural networks and facial recognition to assess the effectiveness of facial feminization

Table 3. Neural Network Confidence in Femininity*

	Average Confidence				
Neural Network and Image View	Male Control	Preoperative FFS	Postoperative FFS	Female Control	*p* (Preoperative vs. Postoperative)
Amazon					
Frontal	−0.74	0.35	0.79	0.72	0.02
Lateral	−0.92	0.19	0.83	0.88	<0.001
IBM					
Frontal	−0.99	0.25	0.98	0.95	<0.001
Lateral	−0.99	0.31	0.87	0.99	0.014
Mean					
Frontal	−0.87	0.30	0.89	0.84	<0.001
Lateral	−0.96	0.25	0.85	0.94	<0.001

FFS, facial feminization surgery.
*Values range from −1.00 to 1.00 (with −1.00 being not confident in femininity/confident in masculinity and +1.00 being confident in femininity).

surgery in a quantitative and objective way. Because deep neural networks are trained on tens of thousands to millions of images,[10] the accuracy of these networks can exceed that of human observers.[11,12] Neural networks algorithmically and autonomously determine which image patterns reflect a face, its gender, mood, and more. For the transgender woman, facial feminization surgery procedures are aimed at bringing facial appearance more in line with female facial features or gender-typical patterns, as recognized by the neural networks.

Our study shows that neural networks were highly accurate at determining the gender of our male and cis-female control subjects. This gender identification accuracy was consistent even with variation in age and ethnicity. Despite significant effort by preoperative facial feminization surgery patients to be identified as female (with hormonal therapy, makeup, and hair), neural network programs still identified them as male almost half of the time. By contrast, postoperative facial feminization surgery patients were recognized as female more than 93 percent of the time. In addition, this was done with significantly greater confidence in femininity.

We were able to realize our hypothesis that neural networking software can be used to show improved gendering after facial feminization surgery. This is one aspect to successful facial feminization surgery. However, there are limitations to our study. First, our study could be improved by including more patients to represent the broad spectrum of preoperative facial feminization surgery patients who undergo different procedures. Moreover, our patients underwent multiple procedures and completed their transitions; future studies could examine the impact of individual procedures along the way to determine which operations produce the greatest effects. Second, more image views of each patient (e.g., oblique or casual "everyday" photographs) rather than just a frontal or lateral view may offer more representative views of what the public is seeing. One of the neural networks (Microsoft) was trained to recognize faces only from a frontal view; as a result, it could not recognize lateral facial views and thus was unable to assign a gender to the lateral views. Third, there is some variability in facial recognition software; a recent study from the MIT Media Lab found that Amazon's Rekognition software misclassified the gender of some darker skinned faces but not lighter skinned faces.[13] Several corporations are now actively working to minimize errors in facial recognition based on biased training data images. It is also worth noting that artificial intelligence in facial identification is already present in many of our lives, and in the future it is likely to be used more often, and perhaps less transparently, by government and corporate interests. Fourth, although the neural networking showed improved gender identification following facial feminization surgery procedures, this may or may not correlate with public gender recognition. A future study will be conducted in which a large, public survey is performed to determine whether actual people (not artificial intelligence) show improvement in gendering following facial feminization surgery. Similar methodology has been used in recent studies for other plastic surgery fields and procedures.[14–16] Finally, as mentioned earlier, there remains a need for larger studies using standardized patient-reported outcome measures, such as the FACE-Q,[17,18] as patient self-perception remains vitally important.[19] We theorize that there will be a strong correlation with high FACE-Q scores and improved gender typing by the public following facial feminization surgery.

CONCLUSIONS

In summary, facial feminization surgery objectively improves the likelihood that transgender women will be recognized as female, as demonstrated by multiple facial recognition neural networks. Although it may be impossible to eliminate all gender mistyping, facial feminization surgery results in a clear, significant, and objective improvement in gender recognition. Patients may therefore be counseled regarding the efficacy and success of facial feminization surgery.

James P. Bradley, M.D.
1991 Marcus Avenue, Suite 102
Lake Success, N.Y. 10042
jpbradley4@mac.com
Instagram: JamesPBradleyMD
Twitter: @DrJamesPBradley
Facebook: @DrJamesPBradley

PATIENT CONSENT

Patients provided written consent for the use of their images.

REFERENCES

1. Ousterhout DK. Feminization of the forehead: Contour changing to improve female aesthetics. *Plast Reconstr Surg.* 1987;79:701–713.
2. Gray MS, Lawrence DT, Golomb BA, Sejnowski TJ. A perceptron reveals the face of sex. *Neural Computation.* 1995;7:1160–1164. DOI:10.1162/neco.1995.7.6.1160.
3. Deschamps-Braly JC. Facial gender confirmation surgery: Facial feminization surgery and facial masculinization surgery. *Clin Plast Surg.* 2018;45:323–331.
4. Gurovich Y, Hanani Y, Bar O, et al. Identifying facial phenotypes of genetic disorders using deep learning. *Nat Med.* 2019;25:60–64.
5. Raffaini M, Magri AS, Agostini T. Full facial feminization surgery: Patient satisfaction assessment based on 180 procedures involving 33 consecutive patients. *Plast Reconstr Surg.* 2016;137:438–448.
6. Ainsworth TA, Spiegel JH. Quality of life of individuals with and without facial feminization surgery or gender reassignment surgery. *Qual Life Res.* 2010;19:1019–1024.
7. Morrison SD, Vyas KS, Motakef S, et al. Facial feminization: Systematic review of the literature. *Plast Reconstr Surg.* 2016;137:1759–1770.
8. Andréasson M, Georgas K, Elander A, Selvaggi G. Patient-reported outcome measures used in gender confirmation surgery: A systematic review. *Plast Reconstr Surg.* 2018;141:1026–1039.
9. Barone M, Cogliandro A, Di Stefano N, Tambone V, Persichetti P. A systematic review of patient-reported outcome measures following transsexual surgery. *Aesthetic Plast Surg.* 2017;41:700–713.
10. Samek W, Binder A, Lapuschkin S, Muller K-R. Understanding and comparing deep neural networks for age and gender classification. In: *2017 IEEE International Conference on Computer Vision Workshops (ICCVW).* Available at: https://ieeexplore.ieee.org/document/8265401. Piscataway, NJ: IEEE; 2017:1629–1638.
11. Huang C, Pan W, Lin S. Gender recognition with face images based on PARCONE mode. In: *Proceedings of the Second Symposium on International Computer Science and Computational Technology (ISCSCT).* 2009:222–226.
12. Wang Y, Kosinski M. Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. *J Pers Soc Psychol.* 2018;114:246–257.
13. Raji ID, Buolamwini J. Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial AI products. In: *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society—AIES 19.* January 2019. DOI:10.1145/3306618.3314244.
14. Lu SM, Hsu DT, Perry AD, et al. The public face of rhinoplasty: Impact on perceived attractiveness and personality. *Plast Reconstr Surg.* 2018;142:881–887.
15. Tse RW, Oh E, Gruss JS, Hopper RA, Birgfeld CB. Crowdsourcing as a novel method to evaluate aesthetic outcomes of treatment for unilateral cleft lip. *Plast Reconstr Surg.* 2016;138:864–874.
16. van Veldhuisen CL, Kamali P, Wu W, et al. Prospective, double-blind evaluation of umbilicoplasty techniques using conventional and crowdsourcing methods. *Plast Reconstr Surg.* 2017;140:1151–1162.
17. Verbruggen C, Weigert R, Corre P, Casoli V, Bondaz M. Development of the facial feminization surgery patient's satisfaction questionnaire (QESFF1): Qualitative phase (in French). *Ann Chir Plast Esthet.* 2018;63:205–214.
18. Sinno S, Schwitzer J, Anzai L, Thorne CH. Face-lift satisfaction using the FACE-Q. *Plast Reconstr Surg.* 2015;136:239–242.
19. Lu SM, Bartlett SP. On facial asymmetry and self-perception. *Plast Reconstr Surg.* 2014;133:873e–881e.