

Construct Validity of the GENDER-Q Face and Neck Scales in 4890 Transgender and Gender Diverse Adults

Anne F. Klassen, DPhil[✉]; Manraj N. Kaur, PhD; Charlene Rae, PhD[✉]; Caleb Haley, MD, MSc[✉]; Fermín Capitán-Cañadas, PhD; Daniel Simon, DMD; Luis Capitán, MD, PhD; Jens U. Berli, MD, MBA; Tim C. van de Grift, PhD, MD, MBA; Margriet G. Mullender, PhD, MBA[✉]; Thomas Satterwhite, MD; Gerhard S. Munding, MD, MBA; Gabriel A. Del Corral, MD; Christodoulos Kaoutzanis, MD; Alireza Hamidian Jahromi, MD; Michael Alperovich, MD, MSc; Fan Liang, MD[✉]; Danny A. Young-Afat, MD, PhD; Sylvie D. Cornacchi, MSc; Andrea L. Pusic, MD, MSc; and Shane D. Morrison, MD, MS

Aesthetic Surgery Journal Open Forum

2026, ojab157

Accepted date: 14 July 2026; online publish-ahead-of-print August 31, 2026.

© The Author(s) 2026. Published by Oxford University Press on behalf of The Aesthetic Society.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com. <https://doi.org/10.1093/asjof/ojab157> www.asjopenforum.com

OXFORD
UNIVERSITY PRESS

Abstract

Background: Gender-affirming facial interventions aim to align facial features of people who identify as transgender or gender diverse (TGD) with their gender identity. GENDER-Q is a comprehensive patient-reported outcome measure with 11 scales that measure satisfaction with the face overall, facial features, specific facial areas, and laryngeal prominence (Adam's apple).

Objectives: The aim of the study was to examine construct validity for the face and neck scales in an international sample of people who identify as TGD.

Methods: An international cross-sectional survey collected GENDER-Q data between February 2022 and March 2024. Participants were aged 18 years and older and had sought or received gender-affirming care within the past 5 years. Predefined hypotheses for expected group-level differences and correlations between GENDER-Q scales were tested using independent *t*-tests, analysis of variance, and Pearson correlation

Dr Klassen is a professor and Dr Rae is a research associate, Department of Pediatrics, McMaster University, Hamilton, Ontario, Canada. Dr Kaur is an investigator, Patient Reported Outcomes, Value and Experience Center, Mass General Brigham, Harvard Medical School, Boston, MA, USA. Dr Haley is a plastic surgery resident physician, Department of Surgery, University of Michigan, Ann Arbor, MI, USA. Dr Capitán-Cañadas is Director of Research and Development and Drs Simon and Capitán are maxillofacial surgeons, The Facialteam Group, HC Marbella International Hospital, Marbella, Málaga, Spain. Dr Berli is Division Chief, Division of Plastic and Reconstructive Surgery, Oregon Health Sciences University, Portland, OR, USA. Dr van de Grift is an assistant professor, Dr Mullender is a professor, and Dr Young-Afat is a plastic surgeon, Department of Plastic, Reconstructive and Hand Surgery, Amsterdam University Medical Center, Amsterdam, The Netherlands. Dr Satterwhite is a plastic surgeon in private practice, San Francisco, CA, USA. Dr Munding is a craniofacial surgeon in private practice, Austin, TX, USA. Dr Del Corral is a plastic surgeon, Department of Plastic and Reconstructive Surgery, MedStar Franklin Square Medical Center, Baltimore, MD, USA. Dr Kaoutzanis is an associate

professor, Department of Surgery, Division of Plastic and Reconstructive Surgery, University of Colorado, Anschutz Medical Campus, Aurora, CO, USA. Dr Hamidian Jahromi is an assistant professor, Department of Plastic and Reconstructive Surgery, Lewis Katz School of Medicine, Temple University, Philadelphia, PA, USA. Dr Alperovich is an associate professor, Department of Surgery, Yale University School of Medicine, New Haven, CT, USA. Dr Liang is an associate professor, Department of Plastic and Reconstructive Surgery, Johns Hopkins University, Baltimore, MD, USA. Mrs Cornacchi is a clinical research coordinator, Department of Pediatrics, McMaster University, Hamilton, Ontario, Canada. Dr Pusic is a Plastic Surgeon, Division of Plastic and Reconstructive Surgery, Mass General Brigham, Boston, MA, USA. Dr Morrison is an associate professor, Division of Plastic Surgery, University of Washington Medical Center, Seattle, WA, USA.

Corresponding Author:

Dr Anne F. Klassen, McMaster University, 1280 Main Street W, Hamilton, ON L8N 3Z5, Canada.

E-mail: aklass@mcmaster.ca

coefficients. The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) guideline was followed to establish construct validity (ie, $\geq 75\%$ hypotheses confirmed).

Results: A total of 4890 participants completed the GENDER-Q Face Overall scale and additional scales as relevant. The sample identified as man (33%), woman (37%), nonbinary (24%), and other gender (6%). The percentage of hypotheses that met the COSMIN criteria of evidence of construct validity ranged from 78% to 100% for the 11 scales. As hypothesized, participants who had undergone gender-affirming facial procedures reported significantly higher satisfaction with appearance scores than those who wanted surgery or required revision surgery.

Conclusions: GENDER-Q scales met COSMIN criteria for hypothesis-based construct validity. GENDER-Q can be used to inform clinical care and to advance research regarding gender-affirming facial interventions.

Level of Evidence: 3 (Therapeutic)

The face plays a central role in how gender is perceived and how people are socially recognized in their daily life. Unlike other aspects of appearance that can be concealed, facial features are always visible and are evaluated by others. As such, the face is closely linked to one's experience of being appropriately gendered. In people who identify as transgender or gender diverse (TGD), incongruence between facial appearance and gender identity can be a source of gender dysphoria and social distress.¹⁻³

Gender-affirming facial interventions aim to align a person's appearance with their gender identity. A range of interventions are available that include minimally invasive procedures (eg, fillers and toxins) and surgery (eg, forehead contouring, jaw and chin shaping, rhinoplasty, and chondrolaryngoplasty).⁴ Such interventions aim to improve gender congruence by aligning facial appearance with one's gender identity.^{1,5} Research has demonstrated that gender-affirming facial interventions are associated with improvements in psychological well-being, patient satisfaction, and health-related quality of life (HRQL).^{1,5-8} As the evidence base for gender-affirming care (GAC) has expanded, facial gender-affirming surgery has gained recognition as medically necessary care rather than a merely cosmetic intervention.^{6,9,10}

Given the inherently subjective nature of facial appearance, outcomes of gender-affirming interventions need to include the patient perspective alongside traditional outcome measures. Patient-reported outcome measures (PROMs) represent indispensable tools for capturing outcomes that matter most to patients. Our team previously developed the FACE-Q Aesthetics PROM to evaluate outcomes of surgical and minimally invasive facial aesthetic procedures from the patient perspective.¹¹⁻¹⁵ This PROM is now widely used in facial aesthetics.¹⁶ However, FACE-Q was developed and validated in cisgender populations and, despite being used to evaluate outcomes in a range of gender-affirming surgical and nonsurgical interventions, it may not adequately capture all relevant concepts unique to TGD people's experiences with GAC.¹⁷⁻²²

To address the need for a PROM specifically for GAC, our team developed the GENDER-Q.²³⁻²⁵ This PROM includes 11 independently functioning scales that measure satisfaction with the face overall, facial features, specific facial areas, and laryngeal prominence (Adam's apple). To inform the development of GENDER-Q scales, cognitive debriefing interviews were conducted with 21 patient participants who provided feedback on the relevance of content from existing FACE-Q Aesthetic scales. Concept elicitation interviews were also conducted with 84 TGD people from 4 countries. Together, the qualitative data were used to create a set of scales that were refined

iteratively with input from 50 experts and 14 patient participants.²⁴

The GENDER-Q international field test study was then conducted, which involved 5497 TGD individuals. Psychometric analyses demonstrated robust content validity, internal consistency, and test-retest reliability of the 11 scales forming the GENDER-Q Face and Neck domain.²⁵ However, assessment of construct validity was lacking.

Construct validity refers to the ability of an instrument to measure its intended construct, which is essential for confirming the instrument's clinical value. The aim of this study was to assess construct validity of the GENDER-Q Face and Neck scales using data from the GENDER-Q field test study. Specifically, we tested a series of a priori hypotheses that examined expected group-level differences and evaluated correlations between GENDER-Q Face and Neck scales and GENDER-Q HRQL scales.

METHODS

Ethics Approval

Development and validation of the GENDER-Q adhered to the established international guidelines and recommendations for PROM development.²⁶⁻²⁸ The cross-sectional GENDER-Q international field test took place between February 2022 and March 2024.²⁵ The present study used a sub-set of participants from the larger GENDER-Q field test who completed the GENDER-Q Face Overall scale and other GENDER-Q scales. Ethics approvals were obtained from the Hamilton Integrated Research Ethics Board (#13171) and participating sites. Participants provided informed consent prior to survey completion.

Data Collection

Detailed data collection methods for the GENDER-Q field test study are described elsewhere.²⁵ Participants were recruited from clinics, social media, Trans PULSE Canada participants and a crowdsourcing platform (ie, Prolific Academic). Data were collected using a REDCap survey with versions in English, Danish, Dutch, and Canadian French.²⁹ Participants were eligible for this study if they were aged 18 years or older, had sought or received GAC within the last 5 years (eg, hormones, surgery, and minimally invasive procedures), and had completed the Face Overall scale. Participants recruited in the first Prolific pilot test completed all Face and Neck scales. After that pilot test, all remaining GENDER-Q participants were asked to complete 4

scales (Face Overall, Facial Features, Upper Face, and Eyebrows), but only participants who wanted to have or had undergone a surgical or minimally invasive procedure for facial features/areas completed the associated scales for the specific parts of the face addressed. For the Adam's Apple scale, in the first Prolific pilot test, participants completed the scale if they wanted to have or had surgery for laryngeal prominence. After the pilot test, for the Adam's Apple scale, participants assigned male sex at birth who wanted or had undergone surgery and anyone who wanted to look more feminine were asked to complete that scale. [Supplemental Figure 1](#) shows recruitment for each scale.

Measures

The face scales measure satisfaction (ie, extremely dissatisfied, very dissatisfied, somewhat dissatisfied, somewhat satisfied, very satisfied, extremely satisfied) with how the face, facial feature, or facial area currently looks. The Adam's Apple scale measures how bothered someone is (extremely, very, somewhat, a little, and not at all) by how their Adam's apple currently looks. Items for the 11 scales cover concepts that patients used to describe their appearance (eg, size, shape, fullness, gender-affirming, attractive, photographs, profile, and up close). To score each scale, the raw total score was converted into a score that ranges from 0 to 100 based on the Rasch-transformed logit scores. Higher scores for the facial scales indicate more satisfied and the Adam's Apple scale indicates being less bothered. Missing scores were imputed using the within-person mean imputation for participants who completed at least 50% of a scale's items. Given that the percentage of cases with missing data in each scale was minimal (2%-4% of cases), and items within the scale measure a similar construct, imputing using the mean was deemed to be appropriate.²⁵ Sensitivity analyses were performed using complete cases to assess stability of results.

Variables used to test hypotheses of expected group-level differences included the following: gender identity, hormone use (ie, type and length of time taking), status of facial interventions (ie, surgery and/or minimally invasive procedures) for each facial area (ie, not sure, don't want, want, had, and had and need more), complications (yes/no), and amount of facial hair (none, a little, a moderate amount, and a lot). Participants were also asked to rate their satisfaction (as above) with the following: height and prominence of their cheekbones, jawline from different angles, size of nose, evenness of nostrils, and shape and fullness of lips. Other items asked participants to rate how often (ie, never, rarely, sometimes, often, and always) they love the way they look, feel affirmed in their gender, and feel attractive. For the Adam's Apple scale, participants were asked how much (ie, not at all, a little, somewhat, and very) they were bothered with its prominence, noticeability, and being misgendered due to it. For correlation hypotheses, the following GENDER-Q scales were used: Gender Dysphoria, Social Acceptance, Psychological Distress, Psychological Well-being, and Sexual Well-being.

Statistical Analysis

Construct validity hypotheses were predefined based on expected group-level differences and correlations between scales, with at least

75% of accepted hypotheses considered sufficient evidence of construct validity according to COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) criteria ([Table 1](#)).²⁶ The hypotheses were developed based on clinical expertise. Expected group-level differences were assessed using independent *T*-tests and analysis of variance (ANOVA) with post-hoc Bonferroni tests or the equivalent nonparametric test. Cohen's *d* and Eta-squared (η^2) were calculated for significant *T*-tests and ANOVAs, respectively. All effect sizes were hypothesized to be at least small in size ($d \geq 0.2$, $\eta^2 \geq 0.01$).^{30,31} Correlations between GENDER-Q scales were interpreted as follows: <0.3 dissimilar constructs, 0.3-<0.5 related but dissimilar constructs, and ≥ 0.5 similar constructs. Normality was assessed both graphically and statistically based on kurtosis (absolute < 7) and skewness (absolute < 2).³² Statistical analyses were performed in IBM SPSS Statistics for Windows, version 30 (IBM Corp., Armonk, NY), with statistical significance $P \leq .05$. The focus of the construct validity analysis was on the consistency of the overall results based on the direction magnitude and statistical significance of the tests, rather than the individual *P*-values.

RESULTS

Demographic and Clinical Characteristics

[Supplemental Table 1](#) shows participant demographic and clinical characteristics for the 4890 GENDER-Q field test participants who completed the Face Overall scale. Participants ranged from 18 to 83 years of age (mean = 32.8; SD = 12.3). Most participants resided in the United States (44%) and Canada (32%), were White (78%), and had a college/university degree or higher (57%). [Supplemental Table 1](#) also provides the sample demographic and clinical characteristics by scale for the Face and Neck domain; the sample size varied by scale due to branching logic.

Of the 4890 participants, 25% (1225) reported that they had no previous facial procedures and wanted to have at least one gender-affirming intervention (ie, surgery and/or a minimally invasive procedure) to change the appearance of their face. The specific types of interventions participants had received were as follows: 80% (472) had surgery, 5% (29) had a minimally invasive procedure, 14% (84) had both, and 2 participants did not specify the procedure type.

GENDER-Q Face and Neck Scale Scores

[Figure 1](#) shows the GENDER-Q Face and Neck scale mean scores by procedure status. For all 11 scales, participants who wanted to have a procedure to change the appearance of their face or neck scored lower than those who had undergone a procedure. For the Face Overall, Facial Features, and specific facial areas scales, those who had previously undergone interventions, but reported that they wanted more, had lower scores than those who reported that they did not want any more facial interventions.

Construct Validity

[Table 1](#) shows the construct validity hypotheses tested and which hypotheses were accepted for each scale. Overall, construct validity

Table 1. Summary of Hypothesis Testing for Construct Validity of GENDER-Q Face and Neck Scales

Cross-sectional hypotheses	Face overall	Facial features	Upper face	Eyebrows	Cheeks	Nose	Nostrils	Lips	Chin	Jawline	Adam's apple
Group Differences—All hypotheses are estimated to have at least a small effect size (ie, $\eta^2 \geq 0.01$ or $d \geq 0.2$)											
Subgroup: take estrogen, want more feminine face, have not had any facial treatment. Scores higher as longer on estrogen	Yes	Yes	—	—	No	—	—	No	—	—	—
Subgroup: take testosterone, want more masculine face, have not had any facial treatment. Scores higher as longer on testosterone	Yes	Yes	—	—	No	—	—	—	—	No	—
Subgroup: want/had and need more/completed facial treatment. Scores incrementally higher from want to had and need more to complete facial treatment (based on facial part)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scores higher if want fewer facial treatments	Yes	Yes	—	—	—	—	—	—	—	—	—
Subgroup: had/need more facial treatment. Scores higher if had surgery vs had minimally invasive treatment	—	—	—	—	No	—	—	No	—	—	—
Subgroup: had/need more facial treatment. Scores higher if none vs current complications	Yes	Yes	—	—	—	—	—	—	—	—	—
Scores incrementally higher as more satisfied with how high cheekbones look	Yes	Yes	—	—	Yes	—	—	—	—	—	—
Scores incrementally higher as more satisfied with how prominent cheekbones look	Yes	Yes	—	—	Yes	—	—	—	—	—	—
Scores incrementally higher as more satisfied with how jawline looks from different angles	Yes	Yes	—	—	—	—	—	—	—	Yes	—
Scores incrementally higher as more satisfied with the size of the nose in relation to their face	Yes	Yes	—	—	—	Yes	—	—	—	—	—
Scores incrementally higher as more satisfied with how even (well-matched) the nostrils look	—	—	—	—	—	Yes	Yes	—	—	—	—
Scores incrementally higher as more satisfied with the shape of the lips	Yes	Yes	—	—	—	—	—	Yes	—	—	—
Scores incrementally higher as more satisfied with how full the lips look	Yes	Yes	—	—	—	—	—	Yes	—	—	—
Scores incrementally higher as more often love the way they look	Yes	Yes	Yes	—	Yes	Yes	—	Yes	Yes	Yes	—
Scores incrementally higher as more often feel affirmed in their gender	Yes	Yes	Yes	—	Yes	Yes	—	Yes	Yes	Yes	—
Scores incrementally higher as more often feel attractive	Yes	Yes	Yes	—	Yes	Yes	—	Yes	Yes	Yes	—
Scores incrementally lower as Adam's apple is more noticeable	—	—	—	—	—	—	—	—	—	—	Yes
Scores incrementally lower as more bothered by prominence of their Adam's apple	—	—	—	—	—	—	—	—	—	—	Yes
Scores incrementally lower as more bothered by being misgendered due to their Adam's apple	—	—	—	—	—	—	—	—	—	—	Yes
Correlational hypotheses	Face Overall	Facial Features	Upper Face	Eyebrows	Cheeks	Nose	Nostrils	Lips	Chin	Jawline	Adam's Apple
Positive correlation with composite score for satisfaction with facial appearance (>0.5 positive)	Yes	—	—	—	—	—	—	—	—	—	—
Subgroup: I wanted feminine face, negative correlation with more facial hair (0.3-0.5)	Yes	Yes	—	—	—	—	—	—	—	—	—
Positive correlation with Face Overall (≥ 0.5 with face domain scales; Adam's Apple 0.3-0.5)	—	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Positive correlation with Facial Features (≥ 0.5 with face scales; Adam's Apple 0.3-0.5)	Yes	—	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Positive correlation with Upper Face (≥ 0.5)	Yes	Yes	—	Yes	Yes	—	—	—	—	—	—

Table 1. Continued

Correlational hypotheses	Face Overall	Facial Features	Upper Face	Eyebrows	Cheeks	Nose	Nostrils	Lips	Chin	Jawline	Adam's Apple
Positive correlation with Eyebrows (≥ 0.5)	Yes	Yes	Yes	—	—	—	—	—	—	—	—
Positive correlation with Cheeks (≥ 0.5)	Yes	Yes	Yes	—	—	—	—	—	—	—	—
Positive correlation with Nose (≥ 0.5)	Yes	Yes	—	—	—	—	Yes	—	—	—	—
Positive correlation with Nostrils (≥ 0.5)	Yes	Yes	—	—	—	Yes	—	—	—	—	—
Positive correlation with Lips (≥ 0.5)	Yes	Yes	—	—	—	—	—	—	—	—	—
Positive correlation with Chin (≥ 0.5)	Yes	Yes	—	—	—	—	—	—	—	Yes	—
Positive correlation with Jawline (≥ 0.5)	Yes	Yes	—	—	—	—	—	—	Yes	—	—
Positive correlation with Body Image (Face Overall/Facial Features ≥ 0.5 ; rest 0.3-0.5)	Yes	Yes	No	Yes	No	Yes	—	Yes	Yes	No	No
Positive correlation with Gender Dysphoria (0.3-0.5)	No	Yes	Yes	Yes	Yes	Yes	—	Yes	Yes	Yes	Yes
Positive correlation with Social Acceptance (0.3-0.5)	Yes	Yes	Yes	Yes	Yes	Yes	—	Yes	No	Yes	—
Positive correlation with Psychological Distress (0.3-0.5)	No	Yes	Yes	Yes	Yes	Yes	—	Yes	Yes	Yes	—
Positive correlation with Psychological Well-being (0.3-0.5)	No	Yes	Yes	Yes	Yes	No	—	Yes	Yes	Yes	—
Positive correlation with Sexual Well-being (0.3-0.5)	Yes	Yes	Yes	Yes	Yes	No	—	No	No	Yes	—
Number of hypotheses accepted	28	30	13	10	14	13	5	13	11	13	7
Number of hypotheses rejected	3	0	1	0	4	2	0	3	2	2	1
Number tested	31	30	14	10	18	15	5	16	13	15	8
% accepted	90	100	93	100	78	87	100	81	85	87	88

results were stable in the complete-case analysis. The percentage of accepted hypotheses ranged from 78% to 100%, demonstrating sufficient evidence of construct validity for each scale. Detailed results for expected group-level differences hypothesis testing are included in [Supplemental Table 2](#). All expected group-level differences hypotheses were accepted for the Face Overall, Facial Features, Upper Face, Eyebrows, Nose, Nostrils, Chin, and Adam's Apple scales. The number of expected group-level differences hypotheses rejected for the other scales were as follows: 3 for Cheeks, 2 for Lips, and 1 for Jawline. Among individuals without facial interventions, participants who reported estrogen use for 2 years or more did not report higher satisfaction scores for Cheeks ($P = .80$) or Lips ($P = .25$) compared with those who reported <2 years of use. Similarly, participants who reported testosterone use for 2 years or more did not report higher satisfaction with Cheeks ($P = .06$) or Jawline ($P = .35$) compared with those who reported <2 years of use. Additionally, no differences in Cheeks and Lips scale scores were observed between participants who had facial surgery and those who had a minimally invasive procedure ($P = .07$ and $P = .71$, respectively).

Detailed results for correlation hypothesis testing are shown in [Supplemental Table 3](#). All correlational hypotheses were accepted for the Facial Features, Eyebrows, and Nostrils scales. The number of correlational hypotheses rejected for the other scales were as follows: 3 for Face Overall, 1 for Upper Face, 1 for Cheeks, 2 for Nose, 1 for Lips, 2 for Chin, 1 for Jawline and 1 for Adam's Apple. Most hypotheses that were not met were weaker than expected, including correlations between Social Acceptance and the Chin scale; Sexual Well-being and the Nose, Lips, and Chin scales; Psychological

Well-being and the Nose scale; and Body Image and the Adam's Apple scale. Hypotheses about correlation between the Face Overall scale and Gender Dysphoria, Psychological Distress, and Psychological Well-being scales were higher than expected. Correlations between Body Image and the Upper Face, Cheeks, and Jawline scales were also higher than expected. These findings indicated that the constructs were more strongly related than hypothesized.

DISCUSSION

The GENDER-Q Face and Neck scales evidenced strong construct validity in an international sample of people who identified as TGD, meeting COSMIN guidelines for all 11 scales.²⁶ Importantly, participants who underwent facial procedures reported higher satisfaction with appearance scores than participants who wanted but had not yet obtained facial interventions. These results demonstrate the GENDER-Q Face and Neck scales' ability to produce meaningful, patient-focused data about the effects of gender-affirming facial interventions.

Previous systematic reviews of PROMs used in GAC, including facial gender-affirming interventions, have consistently shown that few instruments were developed specifically for TGD populations.³³⁻³⁷ Most PROMs used to evaluate treatment outcomes were originally developed in cisgender populations and measure generic constructs rather than outcomes prioritized by TGD individuals. In a review of peer-reviewed literature on facial feminization surgery and patient-reported outcomes up to 2019, 21 facial feminization studies were

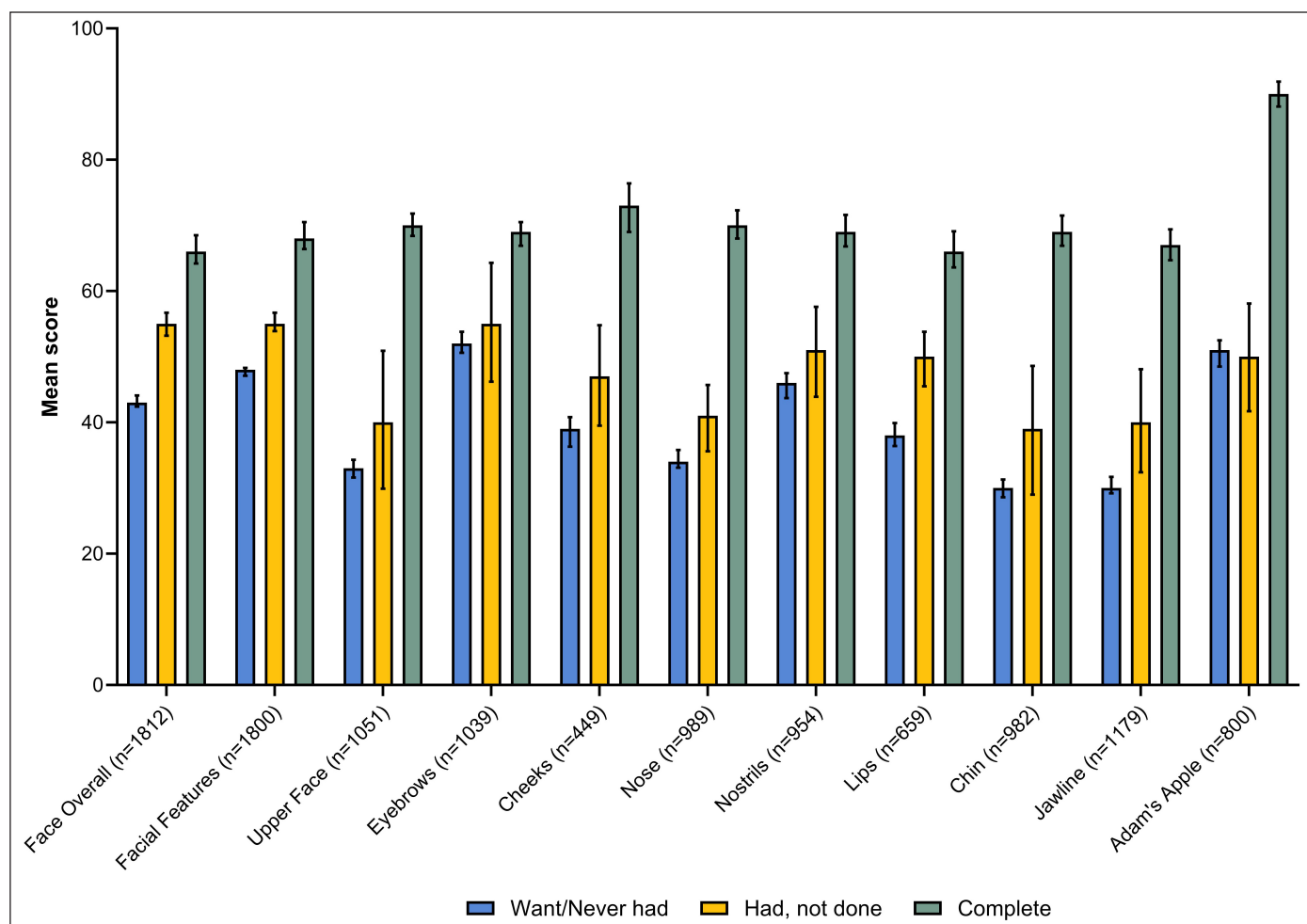


Figure 1. Mean Scores for GENDER-Q Face & Neck scale mean scores by procedure status.

identified.³⁵ This review identified the use of a single TGD-specific PROM (ie, Transgender Congruence Scale). An updated review of literature on facial feminization surgery for TGD patients published in January 2021 found 15 articles which used 12 PROMs.³⁸ None of the 12 instruments used demonstrated strong psychometric properties in patients undergoing facial feminization surgery.

In contrast, the GENDER-Q was developed using extensive input from both clinical experts and TGD adults with lived experience and has been shown to demonstrate high content validity, meaning that it measures relevant outcomes for the TGD population.^{24,25} Since validation of a PROM is an ongoing process with evidence accumulating over time and across multiple studies, the construct validity results presented in this study add to other published development and validation findings.^{24,25} Construct validity is a key measurement property of a PROM as it provides evidence that an instrument measures the theoretical concept it purports to measure.²⁶ With evidence of construct validity, the GENDER-Q Face and Neck scales allow for incorporation of the patient's perspective into outcome assessment for any type of surgical or minimally invasive gender-affirming facial procedure. Such a PROM is needed given the subjective nature of facial appearance and the multi-procedural, region-specific approach often required in gender-affirming facial care.

This study has key implications for the field of facial GAC. Participants who underwent gender-affirming facial procedures reported significantly higher scores than participants who had not yet undergone these procedures. These findings are aligned with published studies that show the benefits associated with this form of care for patients who desire these interventions.¹⁵⁻⁸ Future studies of gender-affirming facial procedures can use the GENDER-Q Face and Neck scales to measure outcomes across gender identities to accurately measure appearance of the face overall, specific facial areas and laryngeal prominence. Potential clinical applications of these scales include using them preoperatively to identify areas of concern for patients. The GENDER-Q can also be used in research to evaluate the effectiveness of different surgical techniques or staging protocols, and to evaluate outcomes of minimally invasive procedures. Importantly, GENDER-Q's modular design provides flexibility for clinicians and researchers to select only the scales relevant to the outcome(s) of interest, thereby minimizing patient burden.

This study has some limitations. The data collected from participants were self-reported and were not verified clinically. Accordingly, some categories used in the analysis may be influenced by factors unrelated to objective treatment outcomes. For example, participants' self-reported intervention status for each facial area (ie, not sure, don't want, want, had, had and need more) may reflect a range of

considerations, including personal preferences, access to care, and treatment goals. The sample was composed mostly of people who are White and from North America. Further research is needed to translate, culturally adapt and validate versions for use in linguistic and cultural contexts beyond the current English, Danish, Dutch, and Canadian French versions. Given the length of the GENDER-Q survey, we were limited in the number of clinical questions we could ask, particularly about types of facial interventions and complications associated with such procedures. Several hypothesized differences and associations were not observed or were weaker in magnitude than anticipated. It is possible these findings relate to the heterogeneity in the sample, and the limited information we were able to collect to understand motivations for facial interventions and revisions. Another limitation is that most of the participants who had facial interventions wanted to look more feminine. Future research could examine construct validity in individuals undergoing facial masculinization.

CONCLUSIONS

The GENDER-Q Face and Neck scales provide a comprehensive set of scales that can be used to accurately measure outcomes of gender-affirming facial interventions, both surgically and minimally invasive, from the patient's perspective. By generating high-quality patient-centered outcomes data related to facial GAC, the GENDER-Q Face and Neck scales can inform patients, providers, and policymakers alike regarding the expected outcomes associated with facial interventions. Future research is needed to support interpretation of scores for GENDER-Q scales, including establishing normative values and determining minimally important differences. Until such benchmarks are established, changes in scores should be interpreted cautiously.

Supplemental Material

This article contains [supplemental material](https://doi.org/10.1093/asjof/ojag157) located online at <https://doi.org/10.1093/asjof/ojag157>.

Acknowledgments

The authors thank each study site, team member, and participant who contributed to the GENDER-Q International Field test.

Disclosures

Drs Kaur, Pusic, and Klassen reported being codevelopers of the GENDER-Q instrument, which is owned by McMaster University and Brigham and Women's Hospital, and receiving a share of license revenues as royalties for its use in for-profit research based on their institution's inventor sharing policy. Dr Kaur reported receiving a Postdoctoral Fellowship Award from the Canadian Institutes of Health Research during the conduct of the study. Dr Fan reported consulting for Integra (Princeton, NJ) outside the submitted work. Dr Alperovich reported consulting for Johnson & Johnson (New Brunswick, NJ) during the conduct of the study. Dr Klassen reported receiving grants from Canadian Institutes of Health Research during the conduct of the study and research consulting through EVENTUM Research (Hamilton, Ontario) outside the submitted work. Dr Haley received National Institutes of Health, grant T32 HS000053. Dr Berli reported owning Jens Berli Consulting (Portland, OR) and providing expert legal counsel and educational consulting to Johnson & Johnson and other

groups outside the submitted work. Dr Kaoutzanis is a consultant for BD (Franklin Lakes, NJ), Bimini Health Tech (Plano, TX), and TELA Bio (Malvern, PA).

Funding

Funding was provided by grant 167288 from the Canadian Institutes of Health Research.

Data availability

Given the personal nature of the data, data will not be publicly available to protect the study participants from the potential risk of identification.

REFERENCES

- Morrison SD, Capitán-Cañadas F, Sánchez-García A, et al. Prospective quality-of-life outcomes after facial feminization surgery: an international multicenter study. *Plast Reconstr Surg*. 2020;145:1499-1509. doi: [10.1097/PRS.0000000000006837](https://doi.org/10.1097/PRS.0000000000006837)
- To M, Zhang Q, Bradlyn A, et al. Visual conformity with affirmed gender or "passing": its distribution and association with depression and anxiety in a cohort of transgender people. *J Sex Med*. 2020;17:2084-2092. doi: [10.1016/j.jsxm.2020.06.006](https://doi.org/10.1016/j.jsxm.2020.06.006)
- Parker AC, Brydges HT, Kimberly LL, Blasdel G, Rodriguez ED. Facing violence: the protective impact of facial gender-affirming surgery. *Plast Reconstr Surg*. 2022;150:934e-945e. doi: [10.1097/PRS.00000000000009351](https://doi.org/10.1097/PRS.00000000000009351)
- Spiegel JH. Facial feminization for the transgender patient. *J Craniofac Surg*. 2019;30:1399-1402. doi: [10.1097/SCS.00000000000005645](https://doi.org/10.1097/SCS.00000000000005645)
- Caprini RM, Oberoi MK, Dejam D, et al. Effect of gender-affirming facial feminization surgery on psychosocial outcomes. *Ann Surg*. 2023;277:e1184-e1190. doi: [10.1097/SLA.00000000000005472](https://doi.org/10.1097/SLA.00000000000005472)
- Raner GA, Jaszkul KM, Bonapace-Potvin M, et al. Quality of life outcomes in patients undergoing facial gender-affirming surgery: a systematic review and meta-analysis. *Int J Transgend Health*. 2023;25:653-662. doi: [10.1080/26895269.2023.2278736](https://doi.org/10.1080/26895269.2023.2278736)
- 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. doi: [10.1097/01.prs.0000475754.71333.f6](https://doi.org/10.1097/01.prs.0000475754.71333.f6)
- 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. doi: [10.1007/s11366-010-9668-7](https://doi.org/10.1007/s11366-010-9668-7)
- Coon D, Berli J, Oles N, et al. Facial gender surgery: systematic review and evidence-based consensus guidelines from the International Facial Gender Symposium. *Plast Reconstr Surg*. 2022;149:212-224. doi: [10.1097/PRS.00000000000008668](https://doi.org/10.1097/PRS.00000000000008668)
- Coleman E, Radix AE, Bouman WP, et al. Standards of care for the health of transgender and gender diverse people, version 8. *Int J Transgend Health*. 2022;23:S1-S259. doi: [10.1080/26895269.2022.2100644](https://doi.org/10.1080/26895269.2022.2100644)
- Klassen AF, Cano SJ, Scott A, Snell L, Pusic AL. Measuring patient-reported outcomes in facial aesthetic patients: development of the FACE-Q. *Facial Plast Surg*. 2010;26:303-309. doi: [10.1055/s-0030-1252030](https://doi.org/10.1055/s-0030-1252030)
- Pusic A, Klassen AF, Scott AM, Cano SJ. Development and psychometric evaluation of the FACE-Q satisfaction with appearance scale: a new patient-reported outcome instrument for facial aesthetics patients. *Clin Plast Surg*. 2013;40:249-260. doi: [10.1016/j.cps.2013.04.004](https://doi.org/10.1016/j.cps.2013.04.004)
- Klassen AF, Cano SJ, Scott AM, Pusic AL. Measuring outcomes that matter to face-lift patients: development and validation of FACE-Q appearance appraisal scales and adverse effects checklist for the lower face and neck. *Plast Reconstr Surg*. 2014;133:21-30. doi: [10.1097/PRS.0000000000000011](https://doi.org/10.1097/PRS.0000000000000011)
- Klassen AF, Cano SJ, East CA, et al. Development and psychometric evaluation of the FACE-Q scales for patients undergoing rhinoplasty. *JAMA Facial Plast Surg*. 2016;18:27-35. doi: [10.1001/jamafacial.2015.0936](https://doi.org/10.1001/jamafacial.2015.0936)
- Klassen AF, Cano SJ, Schwitzer JA, et al. Development and psychometric validation of the FACE-Q skin, lips, and facial rhytides appearance scales and adverse effects checklists for cosmetic procedures. *JAMA Dermatol*. 2016;152:443-451. doi: [10.1001/jamadermatol.2015.4432](https://doi.org/10.1001/jamadermatol.2015.4432)

16. Gallo L, Kim P, Yuan M, et al. Measuring the impact of surgical and nonsurgical facial cosmetic interventions using FACE-Q aesthetic module scales: a systematic review and meta-analysis. *Plast Surg (Oakv)*. 2024;33:403-418. doi: [10.1177/22925503231225480](https://doi.org/10.1177/22925503231225480)
17. Schmidt M, Ramelli E, Atlan M, Cristofari S. FACE-Q satisfaction following upper third facial gender-affirming surgery using custom bone section guides. *Int J Oral Maxillofac Surg*. 2023;52:696-702. doi: [10.1016/j.ijom.2022.11.007](https://doi.org/10.1016/j.ijom.2022.11.007)
18. Almeida MN, Long AS, Junn AH, et al. FACE-Q satisfaction ratings are higher after facial feminization surgery than hormone replacement therapy alone. *Transgend Health*. 2024;9:436-443. doi: [10.1089/trgh.2022.0209](https://doi.org/10.1089/trgh.2022.0209)
19. Alper DP, Almeida MN, Hu KG, et al. Quantifying facial feminization surgery's impact: focus on patient facial satisfaction. *Plast Reconstr Surg Glob Open*. 2023;11:e5366. doi: [10.1097/GOX.0000000000005366](https://doi.org/10.1097/GOX.0000000000005366)
20. Mookerjee VG, Alper DP, Almeida MN, et al. Quantitative analysis of morphometric changes in feminization rhinoplasty utilizing a standardized forehead-rhinoplasty technique. *Aesthet Surg J Open Forum*. 2023;5:ojad095. doi: [10.1093/asjof/ojad095](https://doi.org/10.1093/asjof/ojad095)
21. Perrillat A, Coiante E, SidAhmed M, Graillon N, Hersant B, Meningaud JP. Conservative approach to facial upper third feminisation: a retrospective study. *J Craniomaxillofac Surg*. 2023;51:98-106. doi: [10.1016/j.jcms.2023.01.020](https://doi.org/10.1016/j.jcms.2023.01.020)
22. Slimani M, Ramelli E, Azoulai L, Atlan M, Lellouch AG, Cristofari S. Transwomen satisfaction after nonsurgical facial feminization with botulinum toxin injection of the masseter: a preliminary FACE-Q study. *Ann Chir Plast Esthet*. 2026;71:186-190. doi: [10.1016/j.anplas.2025.09.003](https://doi.org/10.1016/j.anplas.2025.09.003)
23. Klassen AF, Kaur M, Johnson N, et al. International phase I study protocol to develop a patient-reported outcome measure for adolescents and adults receiving gender-affirming treatments (the GENDER-Q). *BMJ Open*. 2018;8:e25435. doi: [10.1136/bmjopen-2018-025435](https://doi.org/10.1136/bmjopen-2018-025435)
24. Kaur MN, Morrison S, Deibert S, et al. International study to develop a patient-reported outcome measure to evaluate outcomes of gender-affirming care (the GENDER-Q). *J Patient Rep Outcomes*. 2024;8:134. doi: [10.1186/s41687-024-00785-x](https://doi.org/10.1186/s41687-024-00785-x)
25. Kaur MN, Rae C, Morrison S, et al. Development and assessment of a patient-reported outcome instrument for gender-affirming care. *JAMA Netw Open*. 2025;8:e254708. doi: [10.1001/jamanetworkopen.2025.4708](https://doi.org/10.1001/jamanetworkopen.2025.4708)
26. Mokkink LBM, Elsman EBM, Terwee CB. COSMIN guideline for systematic reviews of patient-reported outcome measures version 2.0. *Qual Life Res*. 2024;33:2929-2939. doi: [10.1007/s11136-024-03740-x](https://doi.org/10.1007/s11136-024-03740-x)
27. U.S. Department of Health and Human Services Food and Drug Administration. *Guidance for Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims*. Silver Spring (MD): U.S. Food and Drug Administration; 2009. Accessed April 14, 2026. <https://www.fda.gov/media/77832/download>
28. Reeve BB, Wyrwich KW, Wu AW, et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res*. 2013;22:1889-1905. doi: [10.1007/s11136-012-0344-y](https://doi.org/10.1007/s11136-012-0344-y)
29. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. *J Biomed Inform*. Jul 2019;95:103208. doi: [10.1016/j.jbi.2019.103208](https://doi.org/10.1016/j.jbi.2019.103208)
30. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd edn. Routledge; 2013. doi: [10.4324/9780203771587](https://doi.org/10.4324/9780203771587)
31. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Front Psychol*. 2013;4:863. doi: [10.3389/fpsyg.2013.00863](https://doi.org/10.3389/fpsyg.2013.00863)
32. Kim HY. Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restor Dent Endod*. 2013;38:52-54. doi: [10.5395/rde.2013.38.152](https://doi.org/10.5395/rde.2013.38.152)
33. 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. doi: [10.1097/PRS.0000000000004802](https://doi.org/10.1097/PRS.0000000000004802)
34. Barone M, Cogliandro A, Di Stefano N, Tambone V, Persichetti P. A systematic review of patient-reported outcome measures following transsexual surgery. *Aesthet Plast Surg*. 2017;41:700-713. doi: [10.1007/s00266-017-0808-2](https://doi.org/10.1007/s00266-017-0808-2)
35. Oles N, Darrach H, Landford W, et al. Gender affirming surgery: a comprehensive, systematic review of all peer-reviewed literature and methods of assessing patient-centered outcomes (Part 1: breast/chest, face, and voice). *Ann Surg*. 2022;275:e52-e66. doi: [10.1097/SLA.0000000000005256](https://doi.org/10.1097/SLA.0000000000005256)
36. Oles N, Darrach H, Landford W, et al. Gender affirming surgery: a comprehensive, systematic review of all peer-reviewed literature and methods of assessing patient-centered outcomes (Part 2: genital reconstruction). *Ann Surg*. 2022;275:e67-e74. doi: [10.1097/SLA.0000000000005257](https://doi.org/10.1097/SLA.0000000000005257)
37. Morrison SD, Vyas KS, Motake S, et al. Facial feminization: systematic review of the literature. *Plast Reconstr Surg*. 2016;137:1759-1770. doi: [10.1097/PRS.0000000000002436](https://doi.org/10.1097/PRS.0000000000002436)
38. Uhlman K, Gormley J, Churchill I, et al. Outcomes in facial feminization surgery: a systematic review. *Facial Plast Surg Aesthet Med*. 2024;26:236-244. doi: [10.1089/fpsam.2023.0000](https://doi.org/10.1089/fpsam.2023.0000)

Supplementary Table 1. Demographic and clinical characteristics of the study sample by the GENDER-Q Face & Neck scales

		Full Sample n=4890		Facial Parts n=4844		Upper Face n=4780		Eyebrows n=4748		Jawline n=2216		Chin n=1906		Cheeks n=1310		Lips n=1485		Nose n=1815		Nostrils n=1755		Adam's Apple n=1953	
		n	%	n	%	n	%	n	%	n	%	n	%	N	%	n	%	n	%	n	%	n	%
Language	English	4553	93	4515	93	4455	93	4422	93	2129	96	1834	96	1268	97	1441	97	1740	96	1685	96	1817	93
	Danish	136	3	133	3	129	3	132	3	47	2	35	2	22	2	25	2	39	2	34	2	51	3
	Dutch	101	2	97	2	97	2	96	2	25	1	25	1	13	1	14	1	25	1	25	1	55	3
	Canadian French	100	2	99	2	99	2	98	2	15	1	12	1	7	1	5	0	11	1	11	1	30	2
Age at time of survey	18-24 years	1388	28	1371	28	1349	28	1346	28	668	30	515	27	404	31	440	30	500	28	481	27	253	13
	25-29 years	1096	22	1087	22	1074	23	1069	23	471	21	382	20	231	18	242	16	341	19	334	19	354	18
	30-39 years	1301	27	1289	27	1269	27	1260	27	567	26	513	27	323	25	360	24	481	27	464	26	604	31
	40-49 years	527	11	521	11	516	11	507	11	243	11	228	12	149	11	183	12	215	12	206	12	316	16
	50+ years	578	12	576	12	572	12	566	12	267	12	268	14	203	16	260	18	278	15	270	15	426	22
Gender identity	Man	1634	33	1621	34	1600	34	1589	34	485	22	337	18	210	16	218	15	264	15	261	15	0	0
	Woman	1785	37	1773	37	1749	37	1740	37	1076	49	1057	56	704	54	838	56	1069	59	1024	58	1731	89
	Non-binary	1190	24	1172	24	1157	24	1147	24	529	24	419	22	333	25	355	24	392	22	384	22	157	8
	Indigenous / Other cultural identity	44	1	43	1	43	1	41	1	18	1	16	1	11	1	16	1	14	1	14	1	14	1
	Other	237	5	235	5	231	5	231	5	108	5	77	4	52	4	58	4	76	4	72	4	51	3
Sex assigned at birth	Male	2076	42	2063	43	2037	43	2026	43	1269	57	1240	65	833	64	986	66	1245	69	1195	68	1953	100
	Female	2814	58	2781	57	2743	57	2722	57	947	43	666	35	477	36	499	34	570	31	560	32	0	0
Preference for facial appearance	Masculine	2448	50	2418	50	2389	50	2368	50	751	34	511	27	339	26	354	24	421	23	416	24	3	0
	Feminine	2061	42	2046	42	2019	42	2008	42	1254	57	1222	64	818	63	969	65	1226	68	1174	67	1934	99
	Other	379	8	378	8	370	8	370	8	210	10	172	9	152	12	162	11	168	9	165	9	16	1
Country	United States	2172	44	2156	45	2128	45	2109	44	976	44	821	43	564	43	617	42	738	41	697	40	829	42
	Canada	1543	32	1529	32	1509	32	1502	32	532	24	416	22	244	19	312	21	414	23	409	23	657	34
	United Kingdom	295	6	293	6	288	6	286	6	206	9	198	10	145	11	160	11	194	11	194	11	126	7
	Australia	235	5	230	5	225	5	221	5	98	4	83	4	43	3	58	4	73	4	72	4	87	5
	Denmark	139	3	136	3	132	3	135	3	51	2	38	2	25	2	28	2	43	2	38	2	55	3
	Netherlands	120	3	116	2	116	2	115	2	39	2	42	2	25	2	29	2	41	2	41	2	62	3
	Other	380	7	378	7	376	7	374	7	308	14	302	16	259	20	277	19	308	17	300	17	134	8
	Prefer not to answer	5	<1	5	0	5	0	5	0	5	0	5	0	4	0	4	0	4	0	4	0	3	0
Race	White	3797	78	3756	78	3705	78	3688	78	1707	77	1493	78	1026	78	1183	80	1426	79	1381	79	1576	81
	Black	133	3	133	3	131	3	129	3	55	3	50	3	40	3	44	3	51	3	49	3	25	1
	East Asian	77	1	75	2	74	2	74	2	33	2	23	1	13	1	11	1	25	1	25	1	29	2

	Southeast Asian	42	1	42	1	42	1	41	1	17	1	12	1	10	1	10	1	16	1	16	1	13	1
	Indigenous	35	1	35	1	34	1	34	1	15	1	10	1	8	1	9	1	11	1	11	1	12	1
	Latin American	163	3	162	3	159	3	158	3	107	5	94	5	81	6	82	6	94	5	88	5	59	3
	Middle Eastern	35	1	35	1	33	1	33	1	12	1	9	1	7	1	4	0	9	1	9	1	12	1
	South Asian	28	<1	28	1	28	1	28	1	19	1	16	1	15	1	15	1	13	1	13	1	10	1
	Pacific Islander	6	<1	6	0	6	0	6	0	3	0	3	0	1	0	3	0	4	0	4	0	4	0
	Other	15	<1	15	0	15	0	15	0	8	0	7	0	4	0	3	0	4	0	2	0	9	1
	Multiple races	491	10	489	10	485	10	476	10	210	10	164	9	85	7	101	7	136	8	132	8	169	9
	Prefer not to answer	51	1	51	1	51	1	50	1	22	1	19	1	14	1	15	1	21	1	21	1	25	1
	Missing	17	<1	17	0	17	0	16	0	8	0	6	0	6	1	5	0	5	0	4	0	10	1
In the past 3 months, how difficult was it for you to cover your household expenses and pay your bills?	Not at all difficult	1643	35	1623	35	1601	35	1595	35	665	32	607	33	379	30	447	32	583	34	570	34	777	41
	A little difficult	1204	25	1193	26	1176	26	1169	26	544	26	474	26	332	27	374	26	438	25	427	25	432	23
	Somewhat difficult	992	21	984	21	975	21	966	21	465	22	381	21	278	22	310	22	373	22	368	22	343	18
	Very difficult	504	11	503	11	492	11	485	11	252	12	199	11	147	12	160	11	194	11	192	11	182	10
	Extremely difficult	345	7	344	7	339	7	336	7	186	9	162	9	110	9	124	9	144	8	144	9	143	8
Marital status	Never married	2867	59	2836	59	2798	59	2781	59	1388	63	1178	62	818	63	884	60	1110	61	1078	62	946	49
	Separated / divorced / widowed	563	12	557	12	549	12	542	11	278	13	262	14	184	14	227	15	257	14	249	14	380	20
	Married / Common law	1248	26	1240	26	1227	26	1223	26	455	21	393	21	247	19	311	21	375	21	364	21	539	28
	Other	160	3	158	3	154	3	151	3	69	3	54	3	45	4	47	3	54	3	48	3	60	3
	Prefer not to answer	40	1	41	1	40	1	39	1	20	1	13	1	10	1	11	1	13	1	13	1	20	1
Education	Some high school	189	4	188	4	186	4	185	4	96	4	82	4	60	5	64	4	80	5	78	4	73	4
	Completed high school	605	12	600	13	594	13	590	13	297	14	244	13	175	14	191	13	244	14	235	13	212	11
	Some college or trade school or university	1298	27	1288	27	1267	27	1257	27	621	28	516	28	392	31	431	30	501	28	496	28	464	24
	Completed college or trade school or university	1805	37	1786	37	1762	37	1747	37	779	36	682	36	425	33	504	35	627	35	613	35	766	40
	Some Masters or Doctoral degree	281	6	274	6	272	6	271	6	122	6	106	6	69	5	84	6	99	6	97	6	109	6
	Completed Masters or Doctoral degree	666	14	663	14	654	14	655	14	265	12	240	13	155	12	179	12	230	13	227	13	292	15
	Prefer not to answer	17	<1	16	0	16	0	14	0	9	0	9	1	7	1	6	0	8	0	8	1	10	1
Hormone status	No	862	18	850	18	836	18	834	18	488	22	418	22	374	29	400	27	401	22	394	23	89	5
	Yes	3991	82	3957	82	3908	82	3880	82	1708	78	1471	78	921	71	1069	73	1398	78	1349	77	1847	95
Facial treatment status	Do not want	2317	47	2288	47	2256	47	2241	47	239	11	238	13	237	18	237	16	238	13	236	13	376	19
	Want / never had	1225	25	1216	25	1202	25	1191	25	995	45	866	45	602	46	666	45	812	45	774	44	759	39
	Completed	233	5	233	5	231	5	230	5	143	7	155	8	68	5	95	6	161	9	159	9	219	11
	Had not done	354	7	351	7	347	7	343	7	285	13	302	16	206	16	260	18	292	16	285	16	326	17

	Not sure	761	16	756	16	744	16	743	16	554	25	345	18	197	15	227	15	312	17	301	17	273	14
Scalp advancement surgery	No	4511	97	4486	97	4457	97	4462	97	2033	95	1719	94	1179	93	1328	93	1658	94	1618	93	1693	93
	Yes	138	3	138	3	138	3	138	3	112	5	117	6	83	7	108	8	115	7	114	7	136	7
Treatment to change shape or size of brow bone	No	4414	90	4371	90	4313	90	4286	90	1865	84	1526	80	1086	83	1191	81	1445	80	1398	80	1490	77
	Yes	471	10	468	10	462	10	457	10	346	16	375	20	219	17	289	20	365	20	357	20	458	24
Treatment to change shape or size of cheeks	No	4716	97	4671	97	4608	97	4579	97	2073	94	1755	92	1147	88	1345	91	1673	92	1620	92	1790	92
	Yes	170	3	169	4	168	4	165	4	139	6	147	8	159	12	136	9	138	8	135	8	159	8
Treatment to change shape or size of nose	No	4474	92	4430	92	4371	92	4342	92	1904	86	1573	83	1101	85	1223	83	1426	79	1381	79	1553	80
	Yes	407	8	405	8	400	8	397	8	303	14	324	17	200	15	253	17	380	21	374	21	392	20
Treatment to change shape or size of lips	No	4609	95	4565	95	4501	94	4471	94	1990	90	1666	88	1130	87	1218	83	1585	88	1543	88	1683	87
	Yes	269	5	267	6	267	6	265	6	214	10	228	12	168	13	255	17	218	12	212	12	259	13
Treatment to change shape or size of chin	No	4527	93	4483	93	4421	93	4392	93	1915	87	1552	82	1128	87	1250	85	1529	85	1480	84	1609	83
	Yes	357	7	355	7	353	7	350	7	295	13	348	18	176	14	229	16	280	16	275	16	338	17
Treatment to change shape or size of jaws	No	4569	94	4526	94	4466	94	4437	94	1908	86	1625	85	1150	88	1281	86	1565	86	1510	86	1649	85
	Yes	320	6	317	7	313	7	310	7	307	14	280	15	159	12	203	14	249	14	245	14	303	16
Treatment to reduce Adam's apple (born male)	No	1733	84	1722	84	1699	84	1692	84	1027	82	981	80	671	82	794	82	977	80	949	79	1614	83
	Yes	325	16	323	16	320	16	316	16	224	18	241	20	144	18	174	18	250	20	246	21	322	17

Supplementary Table 2. Expected group-level differences hypotheses tests for construct validity

Scale	Hypothesis	N	Mean	SD	SE	95% CI		p-value/ES
						LB	UB	
Subgroup: take estrogen, want more feminine face, have not had any facial treatment. Scores higher as longer on estrogen								
Face Overall	<2 years	122	45	14.9	1.3	42.0	47.4	<0.001 d=0.411
	2 years or more	1210	52	18.0	0.5	50.7	52.7	
Facial Features	<2 years	121	47	10.4	0.9	45.3	49.1	<0.001 d=0.470
	2 years or more	1200	54	15.3	0.4	53.1	54.8	
Cheeks	<2 years	58	49	16.4	2.2	45.0	53.7	0.80
	2 years or more	384	50	20.3	1.0	48.0	52.1	
Lips	<2 years	61	47	15.5	2.0	42.8	50.8	0.25
	2 years or more	452	45	19.3	0.9	43.3	46.8	
Subgroup: take testosterone, want more masculine face, have not had any facial treatment. Scores higher as longer on testosterone								
Face Overall	<2 years	298	52	16.9	1.0	50.4	54.3	<0.001 d=0.542
	2 years or more	1667	63	19.2	0.5	61.7	63.5	
Facial Features	<2 years	298	58	15.2	0.9	56.5	60.0	<0.001 d=0.417
	2 years or more	1646	66	17.9	0.4	64.7	66.4	
Cheeks	<2 years	48	43	25.9	3.7	35.1	50.2	0.06
	2 years or more	147	51	25.0	2.1	46.6	54.7	
Jawline	<2 years	99	40	18.8	1.9	36.5	44.0	0.35
	2 years or more	424	42	21.8	1.1	40.4	44.5	
Subgroup: want/had/need more facial treatment for related facial areas. Scores incrementally higher from want to had and need more to competed facial treatment								
Face Overall	Want / never had	1225	43	15.1	0.4	42.4	44.1	<0.001 η²=0.217
	Had not done (still want procedure on other face parts)	354	55	16.7	0.9	53.2	56.7	
	Completed	233	66	16.7	1.1	64.2	68.5	
Facial Features	Want / never had	1216	48	11.2	0.3	47.1	48.3	<0.001 η²=0.248
	Had not done (still want procedure on other face parts)	351	55	13.3	0.7	53.9	56.7	
	Completed	233	68	15.9	1.0	66.4	70.5	
Upper Face	Want	589	33	16.5	0.7	31.6	34.3	<0.001 η²=0.523
	Had not done	22	40	23.7	5.1	29.9	50.9	
	Completed	440	70	18.3	0.9	68.4	71.8	
Eyebrows	Want	582	52	19.5	0.8	50.6	53.8	<0.001 η²=0.150
	Had not done	22	55	20.4	4.3	46.2	64.3	
	Completed	435	69	19.1	0.9	66.9	70.5	

Cheeks	Want	290	39	19.8	1.2	36.3	40.8	<0.001 η²=0.364
	Had not done	26	47	19.0	3.7	39.5	54.8	
	Completed	133	73	21.8	1.9	69.0	76.4	
Nose	Want	609	34	17.1	0.7	33.1	35.8	<0.001 η²=0.453
	Had not done	56	41	19.0	2.5	35.6	45.7	
	Completed	324	70	20.0	1.1	68.0	72.3	
Nostrils	Want	580	46	23.3	1.0	43.7	47.5	<0.001 η²=0.187
	Had not done	55	51	25.5	3.4	43.9	57.6	
	Completed	319	69	21.7	1.2	66.8	71.6	
Lips	Want	404	38	17.8	0.9	36.4	39.9	<0.001 η²=0.333
	Had not done	50	50	14.5	2.1	45.5	53.8	
	Completed	205	66	19.6	1.4	63.6	69.1	
Chin	Want	634	30	17.2	0.7	28.6	31.3	<0.001 η²=0.485
	Had not done	32	39	27.3	4.8	29.0	48.6	
	Completed	316	69	20.8	1.2	66.9	71.5	
Jawline	Want	872	30	18.6	0.6	29.2	31.7	<0.001 η²=0.394
	Had not done	36	40	23.2	3.9	32.4	48.1	
	Completed	271	67	19.6	1.2	64.7	69.4	
Adam's apple	Want	477	51	22.6	1.0	48.5	52.5	<0.001 η²=0.465
	Had not done	28	50	21.3	4.0	41.7	58.1	
	Completed	295	90	16.7	1.0	88.1	91.9	
Scores higher if want fewer facial treatments								
Face Overall	None	3466	61	18.2	0.3	60.0	61.2	<0.001 η²=0.174
	1	483	50	15.2	0.7	48.6	51.3	
	2	291	46	13.9	0.8	44.7	47.9	
	3	209	43	15.3	1.1	41.0	45.2	
	4	146	43	14.4	1.2	40.4	45.1	
	5	117	40	13.7	1.3	37.5	42.5	
	6 or more	177	33	15.0	1.1	30.3	34.8	
Facial Features	None	3431	64	17.2	0.3	63.0	64.1	<0.001 η²=0.192
	1	478	54	11.4	0.5	52.8	54.9	
	2	289	51	10.1	0.6	49.6	51.9	
	3	206	48	8.9	0.6	46.4	48.9	
	4	146	46	8.0	0.7	44.3	46.9	

	5	117	43	7.2	0.7	41.4	44.0	
	6 or more	176	37	11.2	0.8	35.2	38.5	
Subgroup: had/need more facial treatment. Scores higher if had surgery vs. had minimally invasive treatment								
Cheeks	Just minimally invasive procedure	38	63	21.4	3.5	55.6	69.7	0.07
	Surgery	120	71	23.7	2.2	66.2	74.8	
Lips	Just minimally invasive procedure	47	62	16.6	2.4	57.1	66.8	0.71
	Surgery	205	63	20.5	1.4	60.3	66.0	
Subgroup: had/need more facial treatment. Scores higher if none vs. current complications								
Face	No, I did not have any complications after facial surgery	451	61	17.6	0.8	59.0	62.2	<0.001 d=0.608
	Yes, I have complications now	35	50	14.3	2.4	45.2	55.0	
Facial Features	No, I did not have any complications after facial surgery	449	62	15.9	0.8	60.5	63.5	<0.001 d=0.740
	Yes, I have complications now	35	50	12.0	2.0	46.3	54.5	
Scores incrementally higher as more satisfied with how high cheekbones look								
Face Overall	Extremely Dissatisfied	77	20	14.9	1.7	17.0	23.8	<0.001 η²=0.409
	Very Dissatisfied	112	36	10.2	1.0	34.3	38.2	
	Somewhat Dissatisfied	229	39	13.3	0.9	37.7	41.2	
	Somewhat Satisfied	456	45	12.1	0.6	44.0	46.2	
	Very Satisfied	299	54	13.3	0.8	52.6	55.6	
	Extremely Satisfied	140	70	20.8	1.8	66.6	73.6	
Facial Features	Extremely Dissatisfied	77	29	15.0	1.7	25.2	32.0	<0.001 η²=0.428
	Very Dissatisfied	112	40	10.2	1.0	38.5	42.3	
	Somewhat Dissatisfied	227	45	9.3	0.6	44.1	46.5	
	Somewhat Satisfied	457	49	8.8	0.4	47.7	49.3	
	Very Satisfied	299	56	10.4	0.6	54.8	57.2	
	Extremely Satisfied	140	71	18.7	1.6	67.6	73.9	
Cheeks	Extremely Dissatisfied	76	6	9.7	1.1	4.1	8.6	<0.001 η²=0.677
	Very Dissatisfied	112	29	10.7	1.0	27.3	31.3	
	Somewhat Dissatisfied	226	39	9.9	0.7	37.9	40.5	
	Somewhat Satisfied	454	51	11.7	0.5	49.6	51.7	
	Very Satisfied	299	66	14.1	0.8	64.2	67.4	
	Extremely Satisfied	140	85	20.6	1.7	81.8	88.7	
Scores incrementally higher as more satisfied with how prominent cheekbones look								
Face Overall	Extremely Dissatisfied	83	22	15.1	1.7	19.1	25.7	<0.001 η²=0.411
	Very Dissatisfied	146	36	10.9	0.9	34.2	37.8	

	Somewhat Dissatisfied	272	42	14.0	0.9	40.1	43.5	
	Somewhat Satisfied	432	46	12.2	0.6	44.5	46.9	
	Very Satisfied	257	54	13.0	0.8	52.9	56.1	
	Extremely Satisfied	125	73	20.3	1.8	69.2	76.4	
Facial Features	Extremely Dissatisfied	83	30	15.0	1.6	27.0	33.6	<0.001 η²=0.448
	Very Dissatisfied	146	41	11.0	0.9	39.4	43.0	
	Somewhat Dissatisfied	270	46	9.0	0.5	45.1	47.2	
	Somewhat Satisfied	433	49	8.6	0.4	48.3	49.9	
	Very Satisfied	257	56	10.7	0.7	54.9	57.5	
	Extremely Satisfied	125	74	17.8	1.6	70.6	76.9	
Cheeks	Extremely Dissatisfied	82	8	10.9	1.2	5.3	10.1	<0.001 η²=0.746
	Very Dissatisfied	144	30	10.2	0.8	27.9	31.2	
	Somewhat Dissatisfied	270	41	10.3	0.6	39.9	42.3	
	Somewhat Satisfied	431	52	10.9	0.5	51.2	53.3	
	Very Satisfied	256	69	11.6	0.7	67.7	70.6	
	Extremely Satisfied	125	89	16.9	1.5	85.9	91.9	
Scores incrementally higher as more satisfied with how jawline looks from different angles								
Face Overall	Extremely Dissatisfied	342	32	14.4	0.8	30.1	33.2	<0.001 η²=0.461
	Very Dissatisfied	405	41	11.3	0.6	39.5	41.6	
	Somewhat Dissatisfied	679	47	10.7	0.4	46.4	48.0	
	Somewhat Satisfied	459	53	12.4	0.6	52.1	54.3	
	Very Satisfied	229	62	12.4	0.8	60.9	64.1	
	Extremely Satisfied	94	82	18.1	1.9	78.7	86.1	
Facial Features	Extremely Dissatisfied	340	40	12.7	0.7	38.9	41.6	<0.001 η²=0.454
	Very Dissatisfied	404	46	8.5	0.4	45.1	46.8	
	Somewhat Dissatisfied	677	50	8.5	0.3	49.6	50.9	
	Somewhat Satisfied	460	55	9.7	0.5	54.1	55.8	
	Very Satisfied	229	63	11.4	0.8	61.8	64.8	
	Extremely Satisfied	94	83	15.4	1.6	80.3	86.6	
Jawline	Extremely Dissatisfied	342	10	12.0	0.6	8.9	11.4	<0.001 η²=0.864
	Very Dissatisfied	405	31	7.8	0.4	29.9	31.5	
	Somewhat Dissatisfied	678	43	7.2	0.3	42.2	43.2	
	Somewhat Satisfied	460	56	7.5	0.4	55.2	56.6	
	Very Satisfied	229	71	7.0	0.5	70.3	72.1	

	Extremely Satisfied	94	96	6.7	0.7	94.8	97.5	
Scores incrementally higher as more satisfied with the size of nose in relation to their face								
Face Overall	Extremely Dissatisfied	224	33	15.2	1.0	31.4	35.4	<0.001 η²=0.297
	Very Dissatisfied	219	42	12.1	0.8	40.5	43.7	
	Somewhat Dissatisfied	358	45	14.2	0.8	43.9	46.8	
	Somewhat Satisfied	454	48	13.1	0.6	46.8	49.2	
	Very Satisfied	351	55	13.9	0.7	53.6	56.5	
	Extremely Satisfied	201	69	20.3	1.4	66.0	71.7	
Facial Features	Extremely Dissatisfied	224	39	13.5	0.9	37.3	40.9	<0.001 η²=0.382
	Very Dissatisfied	218	46	8.4	0.6	44.5	46.8	
	Somewhat Dissatisfied	358	49	10.3	0.5	47.7	49.8	
	Somewhat Satisfied	455	50	8.7	0.4	49.7	51.3	
	Very Satisfied	351	57	10.4	0.6	56.2	58.4	
	Extremely Satisfied	201	72	18.0	1.3	69.7	74.7	
Nose	Extremely Dissatisfied	224	16	13.0	0.9	14.5	17.9	<0.001 η²=0.826
	Very Dissatisfied	218	34	6.5	0.4	32.9	34.6	
	Somewhat Dissatisfied	357	43	7.3	0.4	41.9	43.4	
	Somewhat Satisfied	455	51	7.6	0.4	50.6	52.0	
	Very Satisfied	351	66	9.1	0.5	64.7	66.7	
	Extremely Satisfied	201	90	13.4	0.9	88.3	92.0	
Scores incrementally higher as more satisfied with how even (well-matched) the nostrils look								
Face Overall	Extremely Dissatisfied	91	33	20.7	2.2	28.6	37.2	<0.001 η²=0.181
	Very Dissatisfied	82	41	16.2	1.8	37.2	44.3	
	Somewhat Dissatisfied	200	44	16.2	1.1	41.8	46.3	
	Somewhat Satisfied	567	45	13.6	0.6	44.0	46.2	
	Very Satisfied	518	51	13.6	0.6	49.8	52.2	
	Extremely Satisfied	296	62	20.7	1.2	60.0	64.7	
Facial Features	Extremely Dissatisfied	91	38	17.5	1.8	34.1	41.4	<0.001 η²=0.247
	Very Dissatisfied	82	44	11.0	1.2	41.7	46.5	
	Somewhat Dissatisfied	200	47	12.7	0.9	45.5	49.0	
	Somewhat Satisfied	568	49	9.2	0.4	47.7	49.3	
	Very Satisfied	517	54	10.8	0.5	53.1	55.0	
	Extremely Satisfied	296	65	18.2	1.1	63.4	67.5	
Nose	Extremely Dissatisfied	91	19	21.6	2.3	14.6	23.6	<0.001

	Very Dissatisfied	82	37	20.7	2.3	32.3	41.4	$\eta^2=0.348$
	Somewhat Dissatisfied	200	40	19.1	1.4	37.8	43.1	
	Somewhat Satisfied	568	45	14.4	0.6	43.4	45.8	
	Very Satisfied	517	56	17.1	0.8	54.0	57.0	
	Extremely Satisfied	296	73	23.0	1.3	70.3	75.6	
Nostrils	Extremely Dissatisfied	91	11	16.4	1.7	8.1	14.9	<0.001 $\eta^2=0.670$
	Very Dissatisfied	82	35	16.2	1.8	31.5	38.6	
	Somewhat Dissatisfied	200	41	11.9	0.8	39.5	42.9	
	Somewhat Satisfied	567	50	10.8	0.5	48.7	50.4	
	Very Satisfied	516	66	13.5	0.6	65.3	67.6	
	Extremely Satisfied	296	90	18.1	1.1	87.6	91.8	
Scores incrementally higher as more satisfied with the shape of the lips								
Face Overall	Extremely Dissatisfied	95	25	16.0	1.6	21.6	28.1	<0.001 $\eta^2=0.327$
	Very Dissatisfied	162	38	13.0	1.0	35.7	39.7	
	Somewhat Dissatisfied	294	43	12.7	0.7	41.5	44.4	
	Somewhat Satisfied	472	46	14.0	0.6	45.1	47.6	
	Very Satisfied	290	54	14.0	0.8	52.3	55.5	
	Extremely Satisfied	172	68	20.6	1.6	64.9	71.1	
Facial Features	Extremely Dissatisfied	95	32	16.3	1.7	29.1	35.8	<0.001 $\eta^2=0.400$
	Very Dissatisfied	163	42	9.9	0.8	40.3	43.3	
	Somewhat Dissatisfied	294	46	8.6	0.5	45.3	47.3	
	Somewhat Satisfied	472	49	9.3	0.4	48.6	50.3	
	Very Satisfied	290	56	10.6	0.6	55.0	57.5	
	Extremely Satisfied	172	71	18.3	1.4	68.0	73.5	
Lips	Extremely Dissatisfied	95	13	12.0	1.2	10.1	15.0	<0.001 $\eta^2=0.841$
	Very Dissatisfied	163	33	7.3	0.6	31.4	33.6	
	Somewhat Dissatisfied	293	43	6.8	0.4	41.7	43.3	
	Somewhat Satisfied	471	52	6.5	0.3	51.5	52.7	
	Very Satisfied	290	66	8.3	0.5	65.4	67.3	
	Extremely Satisfied	172	89	12.3	0.9	87.2	90.9	
Scores incrementally higher as more satisfied with how full the lips look								
Face Overall	Extremely Dissatisfied	165	30	16.2	1.3	27.2	32.1	<0.001 $\eta^2=0.303$
	Very Dissatisfied	215	41	12.2	0.8	39.2	42.5	
	Somewhat Dissatisfied	337	46	13.4	0.7	44.5	47.4	

	Somewhat Satisfied	380	47	14.3	0.7	45.7	48.6	
	Very Satisfied	237	54	14.7	1.0	52.1	55.9	
	Extremely Satisfied	147	69	21.0	1.7	65.8	72.7	
Facial Features	Extremely Dissatisfied	166	36	15.3	1.2	33.9	38.6	<0.001 η²=0.387
	Very Dissatisfied	215	44	8.7	0.6	42.8	45.1	
	Somewhat Dissatisfied	337	49	7.9	0.4	47.7	49.4	
	Somewhat Satisfied	380	50	10.0	0.5	49.4	51.4	
	Very Satisfied	237	57	11.6	0.8	55.3	58.3	
	Extremely Satisfied	147	72	18.7	1.5	69.4	75.5	
Lips	Extremely Dissatisfied	166	20	13.5	1.1	18.0	22.2	<0.001 η²=0.842
	Very Dissatisfied	215	37	6.9	0.5	36.5	38.4	
	Somewhat Dissatisfied	336	47	6.4	0.4	45.9	47.3	
	Somewhat Satisfied	379	55	7.1	0.4	54.7	56.2	
	Very Satisfied	237	69	7.6	0.5	67.6	69.6	
	Extremely Satisfied	147	92	9.2	0.8	90.6	93.6	
Scores incrementally higher as more often love the way they look								
Face Overall	Never	618	36	15.1	0.6	34.9	37.3	<0.001 η²=0.460
	Rarely	902	47	11.3	0.4	45.9	47.4	
	Sometimes	1396	55	13.1	0.4	54.1	55.5	
	Often	988	65	14.4	0.5	64.0	65.8	
	Always	573	81	17.5	0.7	79.1	81.9	
Facial Features	Never	617	46	13.2	0.5	45.1	47.2	<0.001 η²=0.320
	Rarely	896	52	11.8	0.4	51.6	53.2	
	Sometimes	1390	57	13.1	0.4	56.7	58.1	
	Often	988	65	15.6	0.5	64.2	66.1	
	Always	571	79	17.9	0.7	77.9	80.8	
Upper Face	Never	611	46	21.8	0.9	44.1	47.6	<0.001 η²=0.244
	Rarely	894	55	18.4	0.6	53.3	55.7	
	Sometimes	1382	62	19.0	0.5	60.8	62.8	
	Often	986	69	19.1	0.6	68.0	70.3	
	Always	569	84	19.2	0.8	82.4	85.6	
Cheeks	Never	304	38	22.3	1.3	35.0	40.1	<0.001 η²=0.217
	Rarely	319	48	18.9	1.1	45.9	50.0	
	Sometimes	323	55	19.3	1.1	52.5	56.7	

	Often	183	63	19.3	1.4	60.2	65.8	
	Always	82	75	21.9	2.4	70.3	79.9	
Nose	Never	368	43	22.5	1.2	40.2	44.8	<0.001 η²=0.098
	Rarely	413	47	19.8	1.0	45.5	49.4	
	Sometimes	505	50	20.5	0.9	48.4	52.0	
	Often	296	56	20.3	1.2	54.0	58.6	
	Always	129	69	24.4	2.1	64.9	73.4	
Lips	Never	325	43	21.6	1.2	41.1	45.8	<0.001 η²=0.129
	Rarely	357	50	18.0	1.0	48.2	52.0	
	Sometimes	377	54	17.9	0.9	51.9	55.5	
	Often	217	59	18.1	1.2	56.2	61.0	
	Always	103	72	21.3	2.1	67.4	75.8	
Chin	Never	382	36	22.6	1.2	33.5	38.0	<0.001 η²=0.145
	Rarely	440	43	20.6	1.0	41.3	45.2	
	Sometimes	520	49	20.7	0.9	47.4	50.9	
	Often	301	55	20.9	1.2	53.1	57.8	
	Always	130	69	26.5	2.3	64.2	73.4	
Jawline	Never	442	31	21.1	1.0	28.9	32.8	<0.001 η²=0.187
	Rarely	506	40	19.1	0.8	37.8	41.2	
	Sometimes	615	45	18.9	0.8	43.4	46.4	
	Often	345	53	19.3	1.0	50.6	54.7	
	Always	144	67	26.2	2.2	63.0	71.6	
Scores incrementally higher as more often feel affirmed in their gender								
Face Overall	Never	226	32	17.6	1.2	29.6	34.2	<0.001 η²=0.384
	Rarely	361	41	12.4	0.7	39.7	42.2	
	Sometimes	1019	47	12.5	0.4	46.2	47.8	
	Often	1794	57	13.9	0.3	56.2	57.5	
	Always	1094	73	18.4	0.6	72.2	74.4	
Facial Features	Never	225	44	15.8	1.1	41.7	45.8	<0.001 η²=0.275
	Rarely	361	48	11.5	0.6	47.0	49.4	
	Sometimes	1013	53	12.2	0.4	51.8	53.3	
	Often	1788	59	13.5	0.3	58.3	59.5	
	Always	1091	73	18.4	0.6	72.3	74.5	
Upper Face	Never	222	42	22.9	1.5	39.0	45.0	<0.001

	Rarely	357	48	19.5	1.0	46.3	50.4	$\eta^2=0.214$
	Sometimes	1007	55	19.2	0.6	54.0	56.4	
	Often	1783	63	19.0	0.5	62.2	64.0	
	Always	1088	78	20.5	0.6	76.4	78.9	
Cheeks	Never	156	37	26.1	2.1	33.1	41.3	<0.001 $\eta^2=0.167$
	Rarely	194	45	18.7	1.3	42.6	47.9	
	Sometimes	346	48	18.4	1.0	45.7	49.6	
	Often	351	56	20.3	1.1	53.8	58.0	
	Always	170	69	22.8	1.7	65.8	72.7	
Nose	Never	166	42	23.9	1.9	38.0	45.3	<0.001 $\eta^2=0.065$
	Rarely	228	45	21.8	1.4	41.8	47.5	
	Sometimes	470	48	17.9	0.8	46.2	49.4	
	Often	574	52	20.7	0.9	50.6	54.0	
	Always	280	61	26.1	1.6	57.6	63.7	
Lips	Never	161	43	24.3	1.9	39.4	46.9	<0.001 $\eta^2=0.091$
	Rarely	213	49	18.5	1.3	46.5	51.5	
	Sometimes	387	50	17.4	0.9	48.5	52.0	
	Often	419	54	18.5	0.9	52.0	55.6	
	Always	206	65	22.1	1.5	62.2	68.3	
Chin	Never	171	33	23.2	1.8	29.8	36.8	<0.001 $\eta^2=0.104$
	Rarely	248	43	20.5	1.3	40.2	45.3	
	Sometimes	496	44	19.9	0.9	42.1	45.6	
	Often	586	50	21.8	0.9	48.2	51.7	
	Always	279	61	26.3	1.6	57.8	64.0	
Jawline	Never	184	30	22.0	1.6	26.8	33.2	<0.001 $\eta^2=0.131$
	Rarely	264	36	20.6	1.3	33.4	38.4	
	Sometimes	591	40	18.7	0.8	38.1	41.1	
	Often	723	46	19.8	0.7	45.0	47.9	
	Always	299	59	25.5	1.5	55.9	61.7	
Scores incrementally higher as more often feel attractive								
Face Overall	Never	519	36	15.6	0.7	34.7	37.3	<0.001 $\eta^2=0.399$
	Rarely	816	46	12.4	0.4	44.9	46.6	
	Sometimes	1391	54	12.7	0.3	53.2	54.5	
	Often	1090	63	15.1	0.5	62.5	64.3	

	Always	688	77	19.2	0.7	75.1	78.0	
Facial Features	Never	517	46	14.2	0.6	45.0	47.5	<0.001 $\eta^2=0.273$
	Rarely	813	52	11.9	0.4	50.7	52.4	
	Sometimes	1383	57	12.6	0.3	56.3	57.6	
	Often	1087	64	16.0	0.5	63.0	64.9	
	Always	686	76	18.7	0.7	74.6	77.4	
Upper Face	Never	514	46	21.9	1.0	43.8	47.6	<0.001 $\eta^2=0.210$
	Rarely	804	54	19.4	0.7	52.3	55.0	
	Sometimes	1380	61	18.5	0.5	59.8	61.8	
	Often	1083	69	19.9	0.6	67.3	69.7	
	Always	683	80	20.7	0.8	78.5	81.6	
Cheeks	Never	232	36	21.0	1.4	33.3	38.8	<0.001 $\eta^2=0.182$
	Rarely	291	48	20.5	1.2	45.5	50.3	
	Sometimes	357	53	19.5	1.0	50.9	54.9	
	Often	231	59	21.1	1.4	56.2	61.7	
	Always	119	70	22.0	2.0	66.5	74.5	
Nose	Never	286	41	22.6	1.3	38.0	43.2	<0.001 $\eta^2=0.093$
	Rarely	389	46	19.1	1.0	43.7	47.5	
	Sometimes	529	51	20.0	0.9	49.3	52.7	
	Often	350	56	21.2	1.1	54.1	58.6	
	Always	174	64	25.9	2.0	59.7	67.4	
Lips	Never	260	42	21.2	1.3	39.7	44.9	<0.001 $\eta^2=0.127$
	Rarely	323	48	18.8	1.0	46.4	50.5	
	Sometimes	407	54	17.7	0.9	51.9	55.4	
	Often	269	58	18.5	1.1	55.6	60.0	
	Always	136	68	21.9	1.9	64.7	72.1	
Chin	Never	301	33	22.5	1.3	30.0	35.1	<0.001 $\eta^2=0.154$
	Rarely	409	43	21.3	1.1	40.5	44.6	
	Sometimes	545	48	19.7	0.8	46.8	50.1	
	Often	357	56	20.8	1.1	53.4	57.7	
	Always	179	64	25.9	1.9	60.1	67.7	
Jawline	Never	351	30	20.8	1.1	27.6	31.9	<0.001 $\eta^2=0.165$
	Rarely	459	38	20.4	1.0	35.7	39.4	
	Sometimes	648	44	18.9	0.7	43.0	45.9	

	Often	412	52	19.1	0.9	50.2	53.9	
	Always	202	60	25.8	1.8	56.5	63.7	
Scores incrementally lower as Adam’s apple is more noticeable								
Adam’s Apple	Not At All Noticeable	773	98	8.6	0.3	97.1	98.3	<0.001 η²=0.516
	A Little Noticeable	747	76	19.9	0.7	74.7	77.5	
	Somewhat Noticeable	254	57	22.7	1.4	53.8	59.4	
	Very Noticeable	97	36	25.8	2.6	30.3	40.8	
Scores incrementally lower as more bothered by prominence of their Adam’s apple								
Adam’s Apple	Not At All Bothered	1097	97	6.4	0.2	97.1	97.8	<0.001 η²=0.902
	A Little Bothered	441	69	9.6	0.5	68.5	70.3	
	Somewhat Bothered	236	49	7.9	0.5	47.5	49.5	
	Very Bothered	104	36	8.4	0.8	34.2	37.5	
	Extremely Bothered	67	13	15.1	1.8	9.2	16.5	
Scores incrementally lower as more bothered by being misgendered due to their Adam’s apple								
Adam’s Apple	Not At All Bothered	1219	94	12.4	0.4	92.9	94.3	<0.001 η²=0.638
	A Little Bothered	210	69	13.9	1.0	66.6	70.4	
	Somewhat Bothered	177	57	16.1	1.2	54.3	59.1	
	Very Bothered	141	49	16.6	1.4	45.9	51.4	
	Extremely Bothered	191	40	27.0	2.0	36.0	43.7	

Eta squared (η^2) interpreted as <0.01 negligible, $0.01 \leq \eta^2 < 0.06$ small, $0.06 \leq \eta^2 < 0.14$ Medium, ≥ 0.14 Large; Cohen's d interpreted as <0.2 negligible, $0.2 \leq d < 0.5$ small, $0.5 \leq d < 0.8$ medium, ≥ 0.8 Large; SD – standard deviation, SE – standard error, CI – confidence interval, LB – lower bound, UB – upper bound

Supplementary Table 3. Correlation hypotheses tests for construct validity

		Face Overall	Facial Features	Upper Face	Eyebrows	Cheeks	Nose	Nostrils	Lips	Chin	Jawline	Adam's Apple
Summation of overall satisfaction of each part of the face	Spearman	0.8										
	n	4839										
Amount of facial hair reported for those who identify as woman	Spearman	-0.3	-0.3									
	n	1605	1599									
Face Overall	r		0.8	0.7	0.6	0.7	0.6	0.5	0.6	0.7	0.7	0.3
	n		4839	4777	4741	876	1265	1218	1021	1323	1423	1299
Facial Features	r	0.8		0.8	0.7	0.7	0.7	0.6	0.7	0.7	0.7	0.3
	n	4839		4778	4741	876	1264	1217	1021	1323	1420	1292
Upper Face	r	0.7	0.8		0.7	0.5						
	n	4777	4778		4727	876						
Eyebrows	r	0.6	0.7	0.7								
	n	4741	4741	4727								
Cheeks	r	0.7	0.7	0.5								
	n	876	876	876								
Nose	r	0.6	0.7					0.7				
	n	1265	1264					1218				
Nostrils	r	0.5	0.6				0.7					
	n	1218	1217				1218					
Lips	r	0.6	0.7									
	n	1021	1021									
Chin	r	0.7	0.7								0.7	
	n	1323	1323								1213	
Jawline	r	0.7	0.7							0.7		
	n	1423	1420							1213		
Adam's Apple	r	0.3	0.3									
	n	1299	1292									
Body Image	r	0.7	0.5	0.5	0.4	0.5	0.4		0.4	0.4	0.5	0.2
	n	1622	1620	1615	1617	797	1182		933	1214	1299	1147
Gender Dysphoria	r	0.5	0.4	0.4	0.3	0.4	0.3		0.3	0.4	0.4	0.3
	n	1631	1628	1624	1624	817	1194		954	1227	1310	1163
Social Acceptance	r	0.4	0.4	0.3	0.3	0.3	0.3		0.3	0.2	0.3	
	n	1671	1668	1663	1664	835	1220		971	1253	1345	

Psychological Distress	r	0.5	0.4	0.4	0.3	0.3	0.3		0.3	0.3	0.3	
	n	1600	1597	1593	1593	790	1166		927	1200	1284	
Psychological Well-being	r	0.5	0.4	0.3	0.3	0.3	0.2		0.3	0.3	0.3	
	n	1635	1632	1627	1628	808	1191		942	1224	1312	
Sexual Well-being	r	0.4	0.4	0.3	0.3	0.3	0.2		0.2	0.2	0.3	
	n	1335	1334	1332	1332	657	965		753	993	1071	

r=correlation; n=number; *red text Indicates correlation was not as hypothesized and therefore rejected ; Sample includes only those that wanted, had, or had and needed more