JAMA Network Open

Original Investigation | Equity, Diversity, and Inclusion

Development and Assessment of a Patient-Reported Outcome Instrument for Gender-Affirming Care

Manraj N. Kaur, PhD; Charlene Rae, PhD; Shane D. Morrison, MD, MS; Alexis Laungani, MD; Pierre Brassard, MD; Margriet G. Mullender, PhD, MBA; Tim C. van de Grift, MD, PhD; Danny A. Young-Afat, MD, PhD; Jens Ahm Sørensen, MD, PhD; Lotte Poulsen, MD, PhD; Sylvie D. Cornacchi, MSc; Jack G. Graesser, BA; Michelle Mistry Igbokwe, MD; Thomas Satterwhite, MD; John H. Pang, MD; Arya A. Akhavan, MD; Allison Hu, MD; Natasha Johnson, MD; Stefan J. Cano, PhD; Kinusan Savard, EDD; Gerhard S. Mundinger, MD, MBA; Fermín Capitán-Cañadas, PhD; Daniel Simon, DMD; Luis Capitán, MD, PhD; Devin Coon, MD, MS; Hilliard T. Brydges, MD; Rachel Bluebond-Langner, MD; Eduardo D. Rodriguez, MD, DDS; Lee C. Zhao, MD; Kathleen A. Armstrong, MD, MSc; Nicola R. Dean, MBChB, PhD; Tamara A. Crittenden, PhD; Zac A. Cannell, BSocWk; Megan Lane, MD, MS; Caleb A. Haley, MD; Jessica Hsu, MD, PhD; Geolani W. Dy, MD; Blair R. Peters, MD; Jens U. Berli, MD, MBA; Christina E. Milano, MD; Christian X. Lava, MS; Kenneth L. Fan, MD; Gabriel A. Del Corral, MD; Christodoulos Kaoutzanis, MD; Nargis Kalia, MPH; Ty Higuchi, MD; Oren Ganor, MD; Sangeeta Subedi, MS; Laura M. Douglass, MD; Alireza Hamidian Jahromi, MD; Helia C. Hosseini, MS; Jacqueline Ihnat, BS; Neil Parikh, BA; Kevin Hu, BS; Michael Alperovich, MD, MSc; Edward C. Ray, MD; Youssef Aref, MPH; Bashar A. Hassan, MD, MPH; Fan Liang, MD; Lily Mundy, MD; Mang L. Chen, MD; Andrea L. Pusic, MD, MSc; Anne F. Klassen, DPhil

Abstract

IMPORTANCE There is an urgent need for a validated gender-affirming care–specific patient-reported outcome measure (PROM).

OBJECTIVE To field test the GENDER-Q, a new PROM for gender-affirming care, in a large, international sample of transgender and gender diverse (TGD) adults and evaluate its psychometric properties.

DESIGN, SETTING, AND PARTICIPANTS This international cross-sectional study was conducted among TGD adults aged 18 years and older who were seeking or had received gender-affirming care within the past 5 years at 21 clinical sites across Canada, the United States, the Netherlands, and Spain; participants were also recruited through community groups (eg, crowdsourcing platform, social media). The study was conducted between February 2022 and March 2024. Participants had to be capable of completing the instrument in English, Danish, Dutch, or French-Canadian. Eligible participants accessed an online REDCap survey to complete sociodemographic questions and questions about gender-affirming care they had received or sought (ie, to look, function, or feel masculine, feminine, gender fluid, or another way).

MAIN OUTCOME AND MEASURES Branching logic was used to assign relevant instrument scales. Rasch measurement theory (RMT) analysis was used to examine the fit of the observed data to the Rasch model for each scale. Test-retest reliability and hypothesis-based construct validity of instrument scales were examined. The hypothesis was that instrument scale scores would increase with better outcomes on corresponding categorical questions.

RESULTS A total of 5497 participants (mean [SD] age, 32.8 [12.3] years; 1837 [33.4%] men; 1307 [23.8%] nonbinary individuals; and 2036 [37.0%] women) completed the field test survey. Participants sought or had the following types of gender-affirming care: 2674 (48.6%) masculinizing, 2271 (41.3%) femininizing, and 552 (10.0%) other. RMT analysis led to the development of 54 unidimensional scales and 2 checklists covering domains of health-related quality of life, sexual, urination, gender practices, voice, hair, face and neck, body, breasts, genital feminization, chest, genital masculinization, and experience of care. Test-retest reliability of the scales (intraclass correlation coefficient [average] >0.70) was demonstrated. Only 1 item (phalloplasty donor flap) had

(continued)

Key Points

Question How does a comprehensive patient-reported outcome measure (PROM) to assess outcomes of gender-affirming care perform in an international sample?

Findings In this cross-sectional study, the instrument was developed following internationally established guidelines for PROM development, with data collected from 5497 transgender and gender-diverse adults. Psychometric analysis resulted in 54 independently functioning scales and 2 checklists; test-retest reliability and construct validity of the instrument were established.

Meaning In this study, a modular, scientifically rigorous internationally validated PROM that can be used to measure outcomes of gender-affirming care in clinical care, research, quality improvement, and regulatory efforts was developed and tested.

+ Supplemental content

Author affiliations and article information are listed at the end of this article.

Abstract (continued)

an ICC less than 0.70. As hypothesized, scores increased incrementally with better associated self-reported categorical responses. For example, among 661 participants who reported poor psychological well-being, the mean (SD) scale score was 45 (18) points; for those who reported excellent psychological well-being, the mean (SD) scale score was 85 (16) points ($P < .001$).

CONCLUSIONS AND RELEVANCE In this cross-sectional study of 5497 TGD adults, the instrument demonstrated reliability and validity. The instrument was validated in an international sample and is designed to collect and compare evidence-based outcome data for gender-affirming care from the patients' perspective.

JAMA Network Open. 2025;8(4):e254708. doi:10.1001/jamanetworkopen.2025.4708

Introduction

Gender-affirming care encompasses social, psychological, behavioral, and medical interventions aimed at affirming gender identity and alleviating gender-related distress.[1] Transgender and gender diverse (TGD) individuals, whose gender identity or expression differs from their sex assigned at birth, represent a growing population globally. Many seek gender-affirming care to harmonize aspects of their lives—such as appearance, emotional well-being, and social interactions—with their gender identity. This care is specialized and multidisciplinary, integrating primary, secondary, and tertiary health care services.

Health-related quality of life (HRQL) is a critical measure of the quality and patient-centeredness of gender-affirming care.[2] HRQL is assessed through patient-reported outcomes (PROs) captured by validated questionnaires known as PRO measures (PROMs).[3] PROs provide insights into health care value and symptom changes, addressing potential discrepancies between patient and health care professional perceptions of care and care outcomes.[4-6] Reflecting the importance of PROs, the National Quality Forum in the United States has begun incorporating PROs into quality metrics, and payers and regulators globally are increasingly interested in using PRO data to assess health care value and quality.[7-14]

In gender-affirming care, PROMs have demonstrated utility in measuring improvements in mental health, gender dysphoria, psychosocial outcomes, sexual well-being, and HRQL as well as reductions in anxiety, depression, and suicidality.[15-18] For gender-affirming surgery, PROMs have been used to assess postoperative outcomes, including urination, sexual, satisfaction with appearance, and care experience.[19-25] However, systematic reviews of PROMs in gender-affirming care have highlighted limitations in the current literature, including reliance on ad hoc measures, use of PROMs validated for cisgender populations, and failure to meet international standards for PROM development.[20,21,26-29] The credibility of PROM-based reporting depends on robust psychometric and practical properties within the relevant clinical population. To address gaps in measuring TGD individuals' outcomes and care experiences, we developed a modular PROM specifically designed for gender-affirming care.

Designing the instrument prioritized creating a PROM grounded in the experiences of individuals seeking gender-affirming care, while ensuring ease of use, comprehensiveness, and international applicability for outcome measurement and benchmarking. The instrument was created using a multistep, mixed-methods approach aligned with established PROM development guidelines.[3,30-33] Step 1 involved concept elicitation interviews with adults seeking gender-affirming care from the United States, Canada, Denmark, and the Netherlands, generating a conceptual framework, items, and preliminary scales. Feedback from cognitive debriefing interviews with 7 to 14 patient participants and written or verbal input from 4 to 37 clinicians (number of participants varied by scale) informed iterative refinement. A pilot field test with 601 English-speaking TGD individuals from 30 countries via a crowdsourcing platform called Prolific Academic further refined the scales.

Detailed methods and results for step 1 are available elsewhere.[34,35] To ensure international relevance, the scales were translated into Danish, Dutch, and French-Canadian using best practices for translation and cultural adaptation.[36,37]

This article reports the findings of step 2 of the instrument's development, an international field test study aimed at identifying the best subset of items to retain in each scale. We also provide an analysis of the psychometric properties of reliability and validity, including construct validity, in TGD adults seeking or receiving gender-affirming care.

Methods

Ethics approval for the study was obtained from the Hamilton Integrated Research Ethics Board and from collaborating sites, as detailed in eTable 1 in Supplement 1. All participants provided informed consent. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guidelines for cross-sectional studies was followed.[38] **Figure 1** illustrates the recruitment workflow for the study.

Participants and Procedures

A cross-sectional sample of TGD individuals aged 18 years or older who sought or received gender-affirming care within the past 5 years and could provide informed consent was recruited from 21 clinical sites in Canada, the United States, the Netherlands, and Spain. Participants completed the instrument in English, Danish, Dutch, or French-Canadian (depending on country of recruitment). Recruitment included clinical site–based outreach via emails, social media, patient portals, face-to-face interactions, and posters as well as community-based methods of a crowdsourcing platform (Prolific Academic), outreach to Trans PULSE Canada project participants,[39] closed TGD-specific social media groups, Trans Pride Australia, Copenhagen Pride, and the developer's website. Those seeking care for variations in sex characteristics were excluded (eg, Turner syndrome).

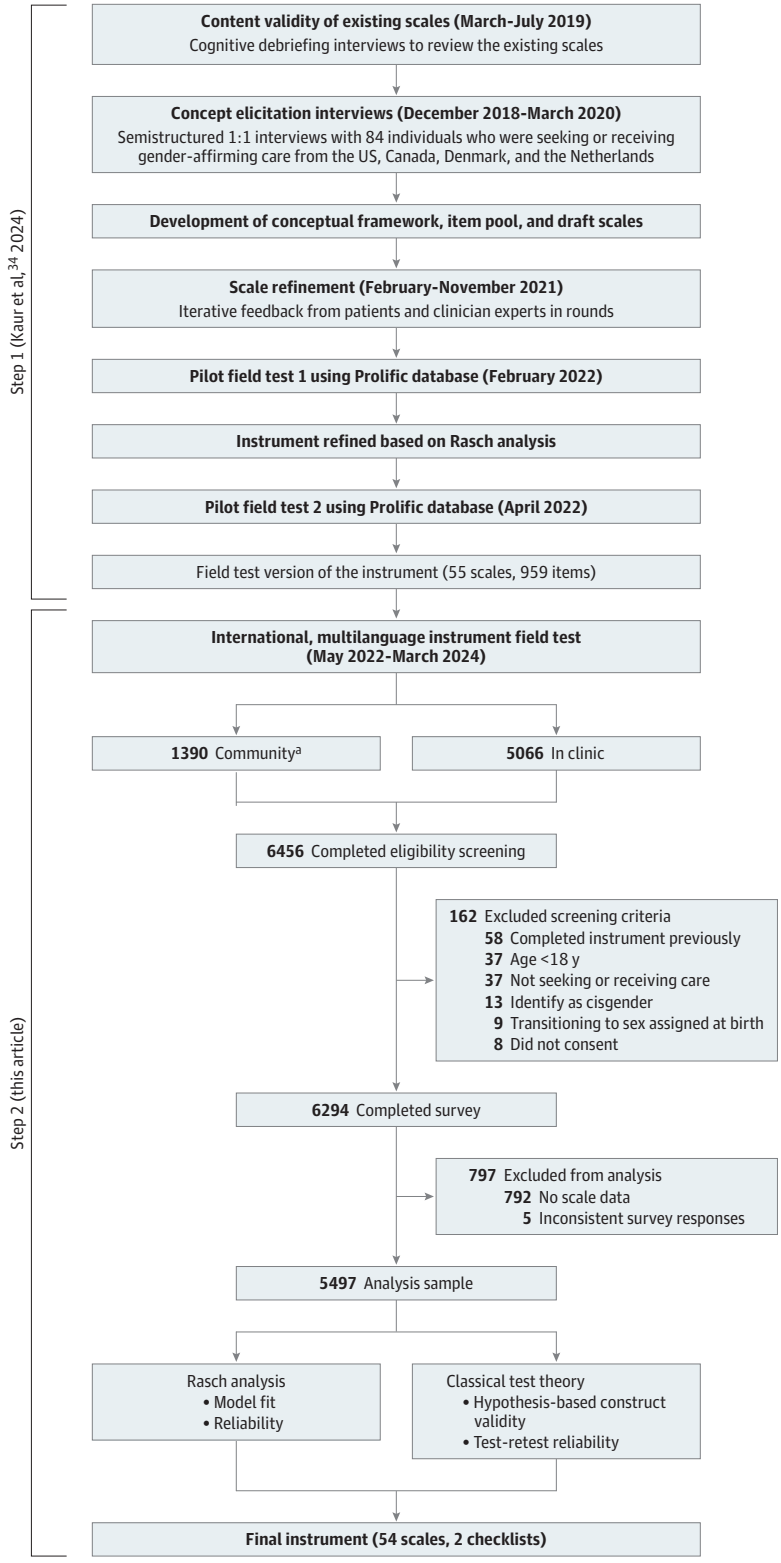
Interested participants completed an online questionnaire hosted on REDCap, with data stored on either McMaster University's or participating sites' servers, based on each site's requirements. A 2-step screener confirmed participants were aged 18 years or older, had sought or received gender-affirming care in the past 5 years, and had not previously completed the survey. Eligible participants provided electronic consent. Field test data collection occurred between February 2022 and March 2024, with the pilot field test conducted from February 2022 to April 2022 and the main field test from May 2022 to March 2024. Participants recruited through the Prolific platform were compensated at a prorated rate of $18 per hour. Those recruited through the Trans PULSE Canada research database received a CAD$25 e-gift card. All other participants, upon completing the survey, could choose to enter a draw for 1 of 10 e-gift cards valued at $100 each, unless prohibited by site institutional regulations.

For test-retest (TRT) reliability, participants recruited through Prolific, Trans PULSE Canada, and 3 clinical sites (GrS Montreal, Crane Center for Transgender Surgery, and GU Recon) who consented to be recontacted for future surveys were invited to complete the scales again. At the start of the TRT survey, participants were queried about any changes in health status, appearance, or the construct being measured by the scale since their initial completion. Data from participants who completed the TRT survey between 7 and 14 days after the initial survey and reported no change were included in the analysis. Participants were compensated with a CAD$25 (or equivalent) e-gift card, and those recruited through Prolific were reimbursed at a prorated rate of $18 per hour.

Measures

Self-reported sociodemographic and clinical data included age, gender identity (open-ended text and check box format), sex assigned at birth, race, education, marital status, sexual orientation, and ability to pay for household expenses and bills. Data on race and ethnicity were collected to describe the sample. Participants indicated the type of gender-affirming care sought or received (to look,

Figure 1. Overview of the Instrument's Development and Field Test



[a] The community sample included data from pilot field tests 1 and 2.

function, or feel masculine, feminine, gender fluid, or none of these). Those identifying as gender fluid or none of these were prompted to specify the type of care by area (eg, face, voice, chest, body, genital). Additional questions covered hormone use, voice surgery, voice therapy, and gender-affirming procedures or operations by body part. For each body part, participants indicated past and future surgical plans using options: I am not sure, I do not want this, I want this, I had this, and I had this and need more, along with the duration of time since treatment. Responses for type of care, sex at birth, and gender-affirming procedures and surgeries used skip logic to ensure only relevant scales were completed. Most demographic and clinical questions included a prefer not to answer option, except where responses were essential for skip logic or analysis.

The PROM Instrument

The field test version of the GENDER-Q comprised of 55 scales and 959 items. Participants completed a core set of scales (n = 12) along with specific scales based on their responses to clinical and sociodemographic variables. Participants had the option to skip items or entire scales if they chose not to respond.

Statistical Analysis

Sociodemographic and clinical characteristics of the sample were summarized using descriptive statistics (continuous variables as means and SDs; categorical variables as numbers and percentages). Data from the pilot and field test were combined for the psychometric analysis. Rasch Measurement Theory (RMT) analysis, a state-of-the-art modern psychometric approach for developing and refining PROMs, was used to examine the fit of the observed data to the Rasch model for each scale.[40] RUMM2030 software (RUMM Laboratory) was used with the unrestricted partial credit model for polytomous data. RMT is grounded in a probabilistic framework that models the relationship between a person's ability or trait level and the difficulty of an item, ensuring that the resulting scales are invariant and interpretable. This approach allows for the construction of unidimensional scales with items that function consistently across groups and provides precise estimates of measurement at both the individual and group level. A series of tests and criteria were applied to determine the optimal subset of items to retain in each scale, aiming to ensure that the scales effectively mapped out a range of measurement for each construct with high reliability and validity (eTable 2 in Supplement 1). Two-sided *P* values of .05 or less were considered significant, with Bonferroni adjustments applied for multiple comparisons where appropriate.

For each scale, items were iteratively examined to determine ordered thresholds, fit to the Rasch model, local dependency, and differential item functioning (DIF). The combined information informed decisions on item retention. For scales with many items exhibiting disordered thresholds, response options were rescored. DIF analyses assessed age groups (18-24, 25-29, 30-39, 40-49, and ≥50 years) and, where applicable, the goal of gender-affirming care (masculinizing or feminizing) to determine whether items functioned consistently across person factors. For scales with more than 500 respondents, the sample size was amended to 500 for tests of fit statistics. To ensure stability, after the final solution for a scale was reached, item fit was reassessed in 5 random samples of 500 participants. To examine reliability, Person Separation Index (PSI) and Cronbach α values were calculated, with reliability values of 0.7 or greater considered sufficient.[41]

Rasch logit scores were transformed to a 0 to 100 scale, where higher scores indicated better outcomes. Classical test theory (CTT) analyses were performed on transformed scores to provide additional evidence of scale performance. Percentage of missing data were computed, based on final item-sets, for each scale. TRT reliability and hypothesis-based construct validity of the scales were examined. For TRT, we aimed to recruit at least 100 participants per scale, the recommended sample size for a very good rating according to COSMIN study design guidelines.[31] A 2-way mixed-effect model evaluating absolute agreement was used to calculate single and average intraclass correlation coefficient (ICC).[42] For each scale, ICC values of 0.70 or greater provided evidence of sufficient reliability.[41] The standard error of measurement and the individual and group level smallest

detectable change (SDC) were also calculated.[43] To assess construct validity, parametric or nonparametric tests were conducted to evaluate the hypothesis that scale scores would increase incrementally with participants reporting better outcomes on the corresponding overall categorical questions for each scale. Categories for the overall questions with fewer than 10 responses in a category were combined for the analysis. The Dilation, Catheter, and Gender Practices scales did not have associated overall questions and were therefore not included in the construct validation analyses. CTT analyses were performed using SPSS version 29 (IBM Corp).

Results

The sample included 5497 participants with a mean (SD) age of 32.8 (12.3) years (range, 18-83 years). Overall, there were 1837 (33.4%) men, 1307 (23.8%) nonbinary individuals, and 2036 (37.0%) women. Sample characteristics are shown in **Table 1**. The type of gender-affirming care participants were seeking or had were 2674 (48.6%) masculinization, 2271 (41.3%) femininization, and 552 (10.0%) other. Most participants were from the US or Canada (4191 [76.2%]), never married (3209 [58.4%]), and had completed college-level education or higher (3071 [55.9%]). Overall, 158 participants (2.9%) were Black, 182 (3.3%) Latin American, and 4236 (77.1%) White. When asked about existing mental or physical health conditions diagnosed by a health professional, 3349 participants (60.9%) reported a mental health condition and 1645 (29.9%) reported a physical health condition. Sample treatment characteristics are shown in **Table 2**. Participants completed a mean (SD) of 22 (8) scales (range, 1-39); mean (SD) scales completed by type of gender-affirming care were: 22 (7) for masculinization; 25 (9), feminization; and 22 (6), other.

Table 3 presents scale-level results. RMT analysis demonstrated the reliability and validity of 52 of 55 scales measuring aspects of HRQL, sexual, urination, gender practices, voice, hair, face and neck, body, breasts, genital feminization, chest, genital masculinization, and experience of care. The 3 exceptions were binding adverse effects, urinary function, and adverse effects of surgery. For the binding scale, several items exhibited disordered thresholds and did not fit the Rasch model. Removing items with disordered thresholds and splitting the scale into 2 scales—binding skin symptoms and binding chest symptoms—resulted in ordered thresholds and satisfied the requirements of the Rasch model. The surgery adverse effects scale also had multiple items with disordered thresholds. Although collapsing the 2 middle response options yielded acceptable item fit statistics, the data did not fit the Rasch model ($P < .05$), and PSI values were only moderate ($\leq$0.7). For the urinary function scale, based on participants who had undergone genital surgery, nearly all items had disordered thresholds. No satisfactory solution was found that ensured ordered thresholds, good item fit, and acceptable reliability. Consequently, the surgical adverse events and urinary function were deemed checklists.

RMT analysis reduced the number of items in the 54 scales by 60.5%, from 904 to 547 items. Items in the treatment outcome, urinary catheter, tucking symptoms, chest scar, vagina, testicular implants, perineum, erectile device, and donor-site adverse events were rescored, resulting in ordered thresholds for all items in the final versions of the scales. For each scale, all items in their final version fit the Rasch model with nonsignificant χ^2 P values after Bonferroni adjustment. Item fit residuals greater than 0.30 were observed for 183 item pairs across 51 scales, indicating some local dependency; however, subtest analyses showed marginal impact on scale reliability ($\leq$0.13 difference in PSI value). None of the PSI values (with and without extremes) dropped below 0.70, except for the clinic scale, where the PSI values with extremes fell to 0.63. For all scales, item coverage was adequate, with no substantial gaps in the measured construct and limited clustering of items. For the 48 scales analyzed for DIF by age group, DIF was detected in 4 items; for the 28 scales analyzed by treatment goal, 47 items showed DIF. When items were split based on DIF variables, Pearson correlations between original and the new split person locations indicated DIF had a negligible impact (all Pearson correlations $\geq$0.95). For the 35 scales completed by more than 500

Table 1. Sociodemographic Characteristics of the Sample

Characteristic	Participants, No. (%) (N = 5497)
Language of the survey	
English	5102 (92.8)
Danish	168 (3.1)
Dutch	118 (2.1)
French Canadian	109 (2.0)
Age at time of survey, y	
18-24	1565 (28.5)
25-29	1230 (22.4)
30-39	1465 (26.7)
40-49	597 (10.9)
≥50	640 (11.6)
Gender identity	
Man	1837 (33.4)
Woman	2036 (37.0)
Nonbinary	1307 (23.8)
Indigenous or other cultural gender minority	49 (0.9)
Another gender	268 (4.9)
Assigned sex at birth	
Male	2363 (43.0)
Female	3134 (57.0)
Goal of the gender-affirming care	
To look, function, or feel more masculine	2673 (48.6)
To look, function, or feel more feminine	2271 (41.3)
To look, function, or feel more gender fluid	481 (8.8)
None of these	69 (1.3)
Missing	3 (0.1)
Intersex	
No	4612 (83.9)
Yes	124 (2.3)
Not sure	704 (12.8)
Prefer not to answer	23 (0.4)
Missing	34 (0.6)
Country of residence	
United States	2440 (44.4)
Canada	1751 (31.9)
United Kingdom	314 (5.7)
Australia	275 (5.0)
Denmark	172 (3.1)
Netherlands	136 (2.5)
Other	402 (7.3)
Prefer not to answer	6 (0.1)
Missing	1 (<0.1)
Race	
Black	158 (2.9)
East Asian	85 (1.5)
Indigenous	43 (0.8)
Latin American	182 (3.3)
Middle Eastern	40 (0.7)
Pacific Islander	6 (0.1)
South Asian	35 (0.6)
Southeast Asian	53 (1.0)
White	4236 (77.1)

(continued)

Table 1. Sociodemographic Characteristics of the Sample (continued)

Characteristic	Participants, No. (%) (N = 5497)
Multiple races	561 (10.2)
Unspecified other or unknown	19 (0.3)
Prefer not to answer	61 (1.1)
Missing	18 (0.3)
Difficulty covering household expenses and paying bills in past 3 mo	
Not at all difficult	1841 (33.5)
A little difficult	1335 (24.3)
Somewhat difficult	1135 (20.6)
Very difficult	574 (10.4)
Extremely difficult	384 (7.0)
Missing	228 (4.1)
Marital status	
Never married	3209 (58.4)
Separated	186 (3.4)
Divorced	419 (7.6)
Widowed	32 (0.6)
Living common-law	479 (8.7)
Married	924 (16.8)
Other	187 (3.4)
Prefer not to answer	48 (0.9)
Missing	13 (0.2)
Education level	
Some high school	215 (3.9)
Completed high school	692 (12.6)
Some college or trade school or university	1469 (26.7)
Completed college or trade school or university	2014 (36.6)
Some master's or doctoral degree	315 (5.7)
Completed master's or doctoral degree	742 (13.5)
Prefer not to answer	19 (0.3)
Missing	31 (0.6)
Mental health condition diagnosed by clinician that is expected to last or has lasted for at least 6 mo	
No	1886 (34.3)
Yes	3349 (60.9)
Prefer not to answer	229 (4.2)
Missing	33 (0.6)
Physical health condition diagnosed by clinician that is expected to last or has lasted for at least 6 mo	
No	3360 (61.1)
Yes	1645 (29.9)
Prefer not to answer	144 (2.6)
Missing	26 (0.5)
Sexual orientation	
Asexual	724 (13.2)
Bisexual	1646 (29.9)
Gay	668 (12.2)
Lesbian	995 (18.1)
Pansexual	1094 (19.9)
Queer	1788 (32.5)
Questioning or unsure	410 (7.5)
Same-gender loving	289 (5.3)
Straight or heterosexual	929 (16.9)
Other	207 (3.8)
Prefer not to answer	50 (0.9)

participants, 5 random samples of 500 participants provided broad support for the final version of the scales.

The scales demonstrated good targeting as the percentage of sample who scored on the outcome scales was high (HRQL, ≥72.5%; sexual, ≥89.3%; urination, ≥92.6%; gender practices, ≥78.1%; voice, ≥89.3%; hair, ≥83.0%; face and neck, ≥70.7; body, ≥87.6%; breast, ≥85.5%; genital feminization, ≥82.7%; chest, ≥69.5%; and genital masculinization, ≥76.4%). The PSI with extremes and without extremes was high (>0.85) for 45 scales and for 49 scales, respectively. For the remaining scales, the PSI was moderate (0.70-0.85), and internal consistency was excellent with Cronbach α values (with and without extremes) of 0.80 or greater.

The CTT analyses showed strong evidence of instrument's reliability and hypothesis-based construct validity. Detailed results for TRT are provided in eTable 3 in Supplement 1 for both single and average ICCs. For TRT, the sample size met COSMIN criteria for a very good rating (≥100) for 34 scales, adequate (50-99) for 1 scale, doubtful (30-49) for 8 scales, and inadequate (<30) for 8 scales. All scales except the phalloplasty donor flap scale (ICC, 0.65; 95% CI, 0.14-0.86) had an ICC (average) of greater than 0.70, ranging from 0.73 (95% CI, 0.30-0.89) for genital masculinization, donor site adverse effects to 0.98 (95% CI, 0.96-0.99) for genital masculinization, penis sensation. Group SDC ranged from 1.0 to 8.4. The catheter and surgery information scales were excluded from the TRT analysis due to the limited number of participants completing these scales (<5). For construct validity, scale scores increased incrementally with better self-reported responses to the overall construct questions. For example, among 661 participants who reported poor psychological well-being, the mean (SD) scale score was 45 (18) points; for those who reported excellent psychological well-being, the mean (SD) scale score was 85 (16) points ($P < .001$) (eTable 4 in Supplement 1). eTables 5 and 6 in Supplement 1 show the proportion of participants to report problems on the items on the urinary function and adverse effects checklist, respectively. eTable 7 in Supplement 1 provides the sample that was used for the analysis for each scale, and the mean scores and key demographic characteristics for the scales.

Table 2. Treatment Characteristics of the Sample

	Participants, No. (%) (N= 5497)	
Type of care	Yes	No
General		
Currently taking hormones	4484 (82.6)	943 (17.4)
Voice		
Voice therapy	1203 (21.9)	4294 (78.1)
Voice surgery	94 (1.7)	5371 (98.3)
Head, face, and neck		
Scalp advancement surgery	139 (3.0)	4556 (97.0)
Surgery or procedure to change shape or size of brow bone	479 (9.6)	4499 (90.4)
Surgery or procedure to change shape or size of nose	415 (8.3)	4559 (91.7)
Surgery or procedure to change shape or size of lips	271 (5.5)	4700 (94.5)
Surgery or procedure to change shape or size of cheeks	172 (3.5)	4807 (96.5)
Surgery or procedure to change shape or size of chin	364 (7.3)	4613 (92.7)
Surgery or procedure to change shape or size of jaws	325 (6.5)	4657 (93.5)
Surgery to reduce Adam's apple	331 (15.7)	1778 (84.3)
Body		
Chest surgery	2295 (76.2)	718 (23.8)
Breast surgery	574 (25.3)	1698 (74.7)
Surgery to change shape or size of waist	168 (3.1)	5186 (96.9)
Surgery to change shape or size of buttocks	93 (1.7)	5262 (98.3)
Genitals		
Surgery to create a penis	420 (13.7)	2655 (86.3)
Surgery to create a vagina	1334 (58.5)	948 (41.5)

Table 3. Scale-Level RMT Results

Scale	Items, No.	Included in RMT, No./total No. (%)	χ^2 (*df*)	*P* value	PSI + E	PSI − E	α + E	α − E	DIF	
									Age	M/F
Health-related quality of life										
Body image	8	4102/4525 (90.7)	76.85 (72)	.33	0.96	0.95	0.97	0.96	0	0
Gender dysphoria	14	4014/4519 (88.8)	96.93 (126)	.98	0.94	0.94	0.97	0.96	0	0
Social acceptance	9	3889/4621 (84.2)	77.06 (63)	.11	0.84	0.84	0.92	0.89	0	0
Psychological distress	10	3881/4454 (87.1)	68.29 (90)	.96	0.90	0.91	0.95	0.93	0	0
Psychological well-being	10	4059/4557 (89.1)	68.66 (90)	.95	0.93	0.92	0.95	0.93	0	0
Treatment outcome	10	2516/3469 (72.5)	108.24 (90)	.09	0.84	0.86	0.94	0.91	0	0
Sexual										
Sexual well-being	12	3674/3898 (94.3)	72.53 (108)	>.99	0.92	0.92	0.94	0.93	1; *r* = 1.0	6; *r* = 1.00
Orgasm	8	1313/1470 (89.3)	78.67 (64)	.10	0.90	0.89	0.93	0.91	0	3; *r* = 1.00
Urination										
Urinary catheter	10	199/225 (92.6)	33.54 (20)	.03	0.87	0.84	0.90	0.86	NC	NC
Gender practices										
Binding, well-being	8	326/367 (88.8)	46.97 (40)	.21	0.93	0.91	0.95	0.93	NC	NC
Binding, chest symptoms	10	322/367 (87.7)	52.11 (40)	.10	0.87	0.87	0.93	0.91	NC	NC
Binding, skin symptoms	5	303/366 (82.8)	39.88 (20)	.01	0.74	0.73	0.85	0.80	NC	NC
Tucking, symptoms	10	239/306 (78.1)	47.42 (30)	.02	0.78	0.80	0.91	0.87	NC	NC
Voice										
Sound	15	5129/5415 (94.7)	85.83 (135)	>.99	0.96	0.96	0.97	0.96	0	7
Distress	10	4794/5367 (89.3)	83.87 (90)	.66	0.92	0.93	0.95	0.95	0	0
Hair										
Hair-face F	7	1315/1584 (83)	84.81 (63)	.04	0.90	0.90	0.95	0.93	0	NA
Hair-face M	12	1887/2043 (92.4)	133.21 (108)	.05	0.95	0.95	0.96	0.96	0	NA
Hair-head	12	1592/1703 (93.5)	108.7 (108)	.46	0.94	0.94	0.96	0.95	0	8
Face and Neck										
Face overall	15	4557/4898 (93.0)	97.89 (135)	.99	0.97	0.97	0.98	0.97	0	0
Facial features	9	4466/4854 (92.0)	61.28 (81)	.95	0.91	0.90	0.93	0.90	0	6
Upper face	9	1441/1582 (91.1)	60.37 (81)	.96	0.96	0.96	0.97	0.96	0	2; *r* = 1.00
Eyebrows	5	1402/1564 (89.6)	37.07 (40)	.60	0.90	0.88	0.93	0.89	0	1
Cheeks	9	773/866 (89.3)	46.12 (54)	.77	0.97	0.96	0.98	0.97	0	0
Nose	10	1293/1438 (89.9)	110.57 (90)	.07	0.96	0.96	0.98	0.96	0	0
Nostrils	7	1164/1355 (85.9)	35.32 (49)	.93	0.96	0.95	0.98	0.96	0	0
Lips	12	975/1050 (92.9)	144.24 (108)	.01	0.97	0.96	0.98	0.97	0	2; *r* = 1.00
Chin	10	1313/1496 (87.8)	111.4 (90)	.06	0.97	0.97	0.99	0.97	0	1; *r* = 1.00
Jawline	10	1698/1905 (89.1)	109.3 (90)	.08	0.97	0.96	0.98	0.97	0	3; *r* = 1.00
Adam's apple	10	810/1146 (70.7)	72.13 (80)	.72	0.94	0.95	0.98	0.96	0	NA
Body										
Body	10	4709/4973 (94.7)	30.8 (90)	>.99	0.95	0.95	0.96	0.96	0	2; *r* = 1.00
Buttocks	10	1260/1406 (89.6)	72.3 (90)	.91	0.96	0.96	0.98	0.97	0	1; *r* = 1.00
Waist	7	1948/2225 (87.6)	43.78 (56)	.88	0.94	0.93	0.96	0.94	0	2; *r* = 1.00
Breasts										
Breasts	12	2041/2131 (95.8)	107.62 (108)	.49	0.93	0.93	0.95	0.94	0	NA
Nipples and areola	8	1777/2071 (85.5)	109.7 (72)	.003	0.91	0.89	0.94	0.91	0	NA
Genital F										
Vagina	10	1106/1236 (89.5)	118.03 (90)	.03	0.89	0.87	0.92	0.89	0	NA
Labia	12	1027/1152 (89.1)	119.67 (108)	.21	0.94	0.93	0.96	0.94	0	NA
Clitoris	6	925/1118 (82.7)	38.69 (36)	.35	0.92	0.90	0.95	0.92	0	NA
Dilation	5	836/930 (89.9)	27.08 (35)	.83	0.89	0.87	0.92	0.88	0	NA

(continued)

Table 3. Scale-Level RMT Results (continued)

Scale	Items, No.	Included in RMT, No./total No. (%)	χ^2 (*df*)	*P* value	PSI + E	PSI − E	α + E	α − E	DIF	
									Age	M/F
Chest										
Chest	10	2186/2857 (76.5)	95.45 (90)	.33	0.94	0.95	0.99	0.98	0	NA
Scars	12	1340/1927 (69.5)	113.51 (108)	.34	0.81	0.85	0.95	0.92	0	NA
Nipples and areola	8	2200/2535 (86.9)	71.4 (72)	.50	0.92	0.91	0.95	0.93	0	NA
Genital M										
Penis	12	366/391 (93.6)	98.74 (60)	.001	0.93	0.93	0.96	0.95	NC	NA
Penis sensation	11	250/281 (89.0)	36.14 (33)	.32	0.95	0.95	0.97	0.95	NC	NA
Glans	9	194/223 (87.0)	19.56 (18)	.36	0.93	0.92	0.96	0.94	NC	NA
Scrotum	10	287/310 (92.6)	70.87 (40)	.002	0.93	0.93	0.95	0.94	NC	NA
Perineum	8	133/174 (76.4)	19.64 (16)	.24	0.90	0.89	0.96	0.92	NC	NA
Donor site, forearm or thigh	8	214/252 (84.9)	34.25 (24)	.08	0.90	0.89	0.95	0.92	NC	NA
Donor site, adverse effects	12	190/251 (75.7)	43.01 (24)	.009	0.70	0.73	0.88	0.84	NC	NA
Testicular implants	10	79/95 (83.2)	12.67 (20)	.89	0.88	0.83	0.93	0.88	NC	NA
Erectile device	12	77/78 (98.7)	22.42 (24)	.55	0.90	0.90	0.93	0.93	NC	NA
Experience of care										
Health professional	15	1308/3017 (43.4)	99.87 (90)	.22	0.74	0.90	0.98	0.97	0	0
Clinic	10	1022/2333 (43.8)	100.71 (70)	.009	0.74	0.88	0.97	0.94	0	0
Surgery, information	10	364/530 (68.7)	57.23 (40)	.04	0.84	0.88	0.95	0.92	NC	NC
Surgery, return to activity	12	280/594 (47.1)	55.4 (48)	.22	0.79	0.91	0.97	0.95	NC	NC

Abbreviations: DIF, differential item functioning; E, extremes; F, feminization; M, masculinization; NA, not applicable; NC, not calculated due to sample size <300; PSI, Person Separation Index; RMT, Rasch Measurement Theory.

Discussion

This instrument is a rigorously developed, modular PROM designed for individuals 18 years and older seeking gender-affirming care, adhering to established PROM development guidelines. It comprises 54 unidimensional scales and 2 checklists, covering a comprehensive range of PROs (**Figure 2**) relevant to gender-affirming care. A mixed-methods approach grounded in extensive input from an international sample of TGD individuals and gender-affirming care clinicians ensured content validity, while psychometric validation confirmed the instrument's reliability and validity.

The instrument addresses the need for a validated, gender-affirming PRO assessment tool for TGD individuals.[20,21,27-29] Unlike existing measures that are developed ad hoc or adapted from cisgender populations, this PROM was developed with TGD individual's input at every stage, including concept elicitation, scale refinement, and international field testing. The use of RMT offers distinct advantages over the commonly used CTT in TGD PROMs, enabling interval-level scoring, improved item function analysis, and generalizability.[44,45] RMT and CTT together robustly established the instrument's reliability and validity for use in gender-affirming care.

To our knowledge, this is the only gender-affirming care–specific PROM developed with a large international sample of TGD individuals. Its modular design allows users to select relevant scales, facilitating integration into clinical care and research while minimizing patient and clinician burden. As the field evolves, new scales can be added to address emerging needs.

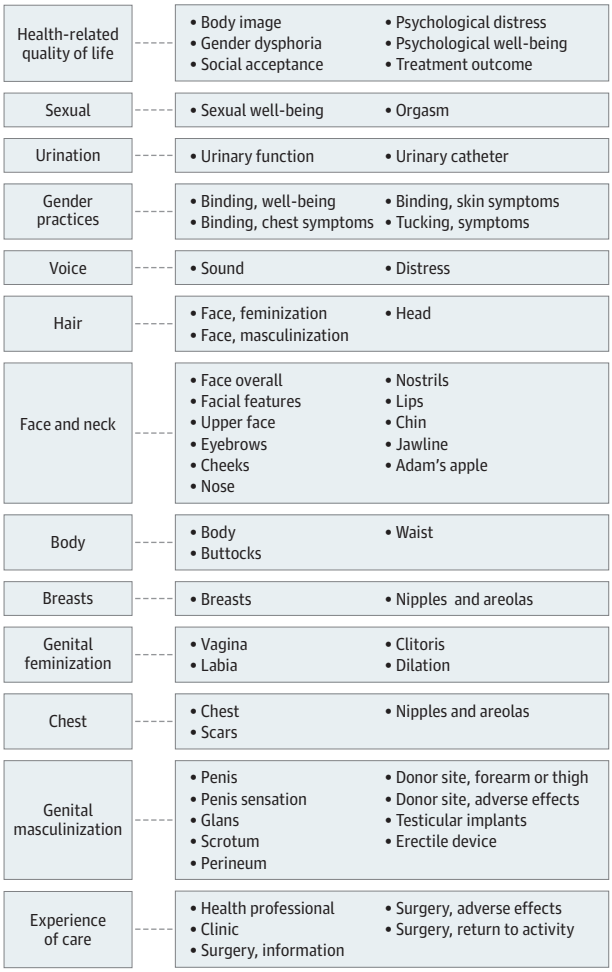
The literature provides examples of how PROMs have been used in health care and social care.[46-49] The PROM described in this article may be used to support clinical care by aiding in screening, risk stratification, expectation management, goal setting, monitoring health status, and facilitating communication between patients and clinicians. Aggregated data could be used to inform care delivery, evaluate interventions, and support health policy decisions, promoting value-based gender-affirming care.

Limitations

The limitations of the field test's design and sample must be considered. The field test relied on self-reported data and online data collection via REDCap, which may have excluded individuals with limited technology skills or access or those living in unsafe or unsupportive environments. Additionally, those unable to engage with lengthy surveys due to time constraints or fatigue may have been excluded, despite the option to save and return later. While the study benefits from an international sample of TGD adults, the predominantly White sample limits generalizability to racially, ethnically, and geographically diverse populations. This demographic pattern reflects broader challenges in diversifying research samples in gender-affirming care, where structural inequities may limit access to both care and research participation. This limitation is particularly relevant as experiences and outcomes of TGD individuals may vary across sociocultural and systemic contexts.

Additionally, small sample sizes constrained testing for some scales. The proportion of participants scoring on the experience of care scales was low (43.4%-68.7%), consistent with literature indicating high ceiling effects in health care experience evaluations.[50-53] For TRT, the sample size met COSMIN criteria for a very good rating (≥100) for 34 scales, adequate for 1 scale, doubtful for 8 scales, and inadequate for 8 scales. Furthermore, the TRT for urinary catheter and information scales were not tested due to small sample size. Although the instrument was validated in 4 languages, establishing the cross-cultural validity of the PROM will be critical to ensuring its robustness and utility in diverse global contexts. Future publications will further examine the

Figure 2. Conceptual Framework of the Patient-Reported Outcome Measure Instrument



construct validity of the scales, while additional studies could assess their reliability in independent clinical samples.

Conclusions

This novel PROM instrument consisting of 54 independently functioning scales and 2 checklists demonstrated reliability and validity in a large international sample of TGD adults. This is a rigorously developed instrument for use in gender-affirming care, research, quality improvement, and regulatory efforts and is available online.[54]

ARTICLE INFORMATION

Accepted for Publication: February 2, 2025.

Published: April 18, 2025. doi:10.1001/jamanetworkopen.2025.4708

Corresponding Author: Manraj N. Kaur, PhD, Patient-Reported Outcomes and Values, & Experience Center (PROVE), Brigham and Women's Hospital, Harvard Medical School, 75 Francis St, Boston, MA 02116 (mkaur17@bwh.harvard.edu).

Author Affiliations: Patient-Reported Outcomes and Values, & Experience Center (PROVE), Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts (Kaur); Department of Pediatrics, McMaster University, Hamilton, Ontario, Canada (Rae, Cornacchi, Johnson, Klassen); University of Washington Medical Center, Seattle (Morrison); Plastic Surgery Division, GrS Montréal, Montreal, Quebec, Canada (Laungani, Brassard); Department of Plastic Reconstructive and Hand Surgery, Amsterdam University Medical Center, Amsterdam, the Netherlands (Mullender, van de Grift, Young-Afat); Research Unit for Plastic Surgery, University of Southern Denmark, Odense, Denmark (Sørensen, Poulsen, Graesser); Department of Plastic Surgery and Burns Treatment, Rigshospitalet, Copenhagen, Denmark (Igbokwe); Align Surgical Associates Inc, San Francisco, California (Satterwhite, Pang, Akhavan, A. Hu); Modus Outcomes (a division of Thread), St James House, Cheltenham, United Kingdom (Cano); Fleming College, Noelville, Ontario, Canada (Savard); Crane Center for Transgender Surgery, Austin, Texas (Mundinger); The Facialteam Group, HC Marbella International Hospital, Málaga, Spain (Capitán-Cañadas, Simon, Capitán); Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts (Coon); Hansjörg Wyss Department of Plastic Surgery, NYU Grossman School of Medicine, New York, New York (Brydges, Bluebond-Langner, Rodriguez); Department of Urology, NYU Grossman School of Medicine, New York, New York (Zhao); Department of Surgery, Women's College Hospital, Toronto, Ontario, Canada (Armstrong); College of Medicine and Public Health, Flinders University, Bedford Park, South Australia, Australia (Dean, Crittenden); Trans Health South Australia, Level 5, Flinders Medical Centre, Bedford Park, South Australia, Australia (Cannell); Section of Plastic Surgery, University of Michigan, Ann Arbor (Lane, Haley, Hsu); Department of Urology, Oregon Health & Science University, Portland (Dy); Department of Surgery, Division of Plastic and Reconstructive Surgery, Oregon Health & Science University, Portland (Peters, Berli); Family Medicine, Oregon Health & Science University, Portland (Milano); Department of Plastic and Reconstructive Surgery, MedStar Georgetown University Hospital, Washington, DC (Lava, Fan); Department of Plastic and Reconstructive Surgery, MedStar Franklin Square Medical Center, Baltimore, Maryland (Del Corral); Department of Surgery, Division of Plastic and Reconstructive Surgery, University of Colorado, Anschutz Medical Campus, Aurora (Kaoutzanis, Kalia); Department of Surgery, Division of Urology, University of Colorado, Anschutz Medical Campus, Aurora (Higuchi); Department of Plastic and Oral Surgery, Boston Children's Hospital, Boston, Massachusetts (Ganor, Subedi); Department of Urology, Lewis Katz School of Medicine at Temple University, Philadelphia, Pennsylvania (Douglass); Department of Plastic and Reconstructive Surgery, Lewis Katz School of Medicine at Temple University, Philadelphia, Pennsylvania (Hamidian Jahromi); Plastic and Reconstructive Surgery, Yale School of Medicine, New Haven, Connecticut (Hosseini, Ihnat, Parikh, K. Hu, Alperovich); Department of Surgery, Cedars-Sinai Medical Center, Los Angeles, California (Ray, Aref); Plastic Surgery, Johns Hopkins, Baltimore, Maryland (Hassan, Liang, Mundy); G.U. Recon, San Francisco, California (Chen); Division of Plastic and Reconstructive Surgery, Brigham and Women's Hospital, Boston, Massachusetts (Pusic).

Author Contributions: Drs Klassen and Kaur had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Kaur, Morrison, van de Grift, Young-Afat, Poulsen, A. Hu, Johnson, Cano, Savard, Coon, Bluebond-Langner, Rodriguez, Hsu, Fan, Ganor, Hassan, Pusic, Klassen.

Acquisition, analysis, or interpretation of data: Kaur, Rae, Morrison, Laungani, Brassard, Mullender, van de Grift, Young-Afat, Sørensen, Cornacchi, Graesser, Igbokwe, Satterwhite, Pang, Akhavan, Mundinger, Capitán-Cañadas, Simon, Capitán, Coon, Brydges, Bluebond-Langner, Zhao, Armstrong, Dean, Crittenden, Cannell, Lane, Haley, Dy, Peters, Berli, Milano, Lava, Fan, Del Corral, Kaoutzanis, Kalia, Higuchi, Ganor, Subedi, Douglass, Hamidian Jahromi, Hosseini, Ihnat, Parikh, K. Hu, Alperovich, Ray, Aref, Liang, Mundy, Chen, Pusic, Klassen.

Drafting of the manuscript: Kaur, Rae, Morrison, Brassard, Young-Afat, Cornacchi, Igbokwe, Savard, Capitán-Cañadas, Simon, Capitán, Coon, Lava, Hosseini, Hassan, Chen, Pusic, Klassen.

Critical review of the manuscript for important intellectual content: Kaur, Rae, Morrison, Laungani, Mullender, van de Grift, Young-Afat, Sørensen, Poulsen, Cornacchi, Graesser, Satterwhite, Pang, Akhavan, A. Hu, Johnson, Cano, Savard, Mundinger, Coon, Brydges, Bluebond-Langner, Rodriguez, Zhao, Armstrong, Dean, Crittenden, Cannell, Lane, Haley, Hsu, Dy, Peters, Berli, Milano, Lava, Fan, Del Corral, Kaoutzanis, Kalia, Higuchi, Ganor, Subedi, Douglass, Hamidian Jahromi, Hosseini, Ihnat, Parikh, K. Hu, Alperovich, Ray, Aref, Hassan, Liang, Mundy, Chen, Pusic, Klassen.

Statistical analysis: Kaur, Rae, Cano, Capitán-Cañadas, Cannell, Klassen.

Obtained funding: Kaur, van de Grift, Pusic, Klassen.

Administrative, technical, or material support: Kaur, Rae, Brassard, Mullender, van de Grift, Young-Afat, Satterwhite, Pang, Akhavan, A. Hu, Johnson, Savard, Mundinger, Capitán-Cañadas, Simon, Capitán, Coon, Brydges, Zhao, Armstrong, Crittenden, Lane, Haley, Dy, Lava, Del Corral, Kalia, Subedi, Douglass, Hamidian Jahromi, Ihnat, Parikh, Liang, Mundy, Klassen.

Supervision: Kaur, Mullender, van de Grift, Young-Afat, Sørensen, Poulsen, Igbokwe, Akhavan, Bluebond-Langner, Rodriguez, Armstrong, Hsu, Lava, Fan, Kaoutzanis, Ganor, Hosseini, Parikh, Aref, Chen, Pusic, Klassen.

Conflict of Interest Disclosures: Drs Kaur, Pusic, and Klassen reported being codevelopers of the GENDER-Q instrument, which is owned by McMaster University and Mass General Brigham, 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 grants from the Canadian Institutes of Health Research during the conduct of the study. Dr Rae reported receiving grants from the Canadian Institutes of Health Research during the conduct of the study. Dr Igbokwe reported receiving grants from the Canadian Institutes of Health Research during the conduct of the study. Dr Cano reported being president of Modus Outcomes outside the submitted work. Dr Dean reported being involved in an application to the Australian government about improving public funding for gender-affirming surgery outside the submitted work; this work has been unpaid. Dr Berli reported owning Jens Berli Consulting and providing expert legal counsel and educational consulting to Johnson & Johnson and other groups outside the submitted work. Dr Fan reported consulting for Integra outside the submitted work. Dr Alperovich reported consulting for Johnson & Johnson during the conduct of the study. Dr Pusic reported receiving grants from the Canadian Institutes of Health Research during the conduct of the study. Dr Klassen reported receiving grants from Canadian Institutes of Health Research during the conduct of the study and consulting for Anne F Klassen Research through EVENTUM Research outside the submitted work. No other disclosures were reported.

Funding/Support: Funding for the field test was provided by grant 167288 from the Canadian Institutes of Health Research.

Role of the Funder/Sponsor: The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Data Sharing Statement: See Supplement 2.

Additional Contributions: We would like to thank the following individuals with their assistance with study conduct: Kamilla Kamaruddin, MD, East of England Gender Service, Cambridge, United Kingdom (data collection); Rixt AC Luikenaar, MD, Rebirth Obgyn/Rebirth Health Center/Department of Obstetrics and Gynecology, University of Utah (data collection); Loren S. Schechter, MD, Department of Surgery, Rush University Medical Center, Chicago, Illinois (data collection); Ian T. Nolan, MD, Department of Surgery, Rush University Medical Center, Chicago, Illinois (data collection); Brielle Weinstein, MD, Department of Surgery, Rush University Medical Center, Chicago, Illinois (data collection); Kristin M. Jacobs, MD, RUSH Gender Affirming Surgery, Rush University Medical Center, Chicago, Illinois (data collection); Sarine Lebrun, BAA, GrS Montreal, Montreal, Quebec, Canada (data collection); Emilie Potts, BSN, GrS Montreal, Montreal, Quebec, Canada (data collection); Kristin C. Daniel, MSN, Crane Center for Transgender Surgery, Austin, Texas (data collection); Morgan B. Menendez, MBA, Crane Center for Transgender Surgery, Austin, Texas (data collection); Curra García, BS (IT manager), The Facialteam Group HC Marbella International Hospital, Malaga, Spain (data collection); Lilia Koss, BA, The Facialteam Group HC Marbella International Hospital, Malaga, Spain (data collection); Augustus Parker, BS, Grossman School of Medicine, New

York University, New York (data collection); Gaines Blasdel, BS, Grossman School of Medicine, New York University, New York (data collection); Paul Mittermiller, MD, Align Surgical Associates Inc, San Francisco, California (data collection); Michael Safir, MD, Align Surgical Associates Inc, San Francisco, California (data collection); Dev Gurjala, MD, Align Surgical Associates Inc, San Francisco, California (data collection); Breanna Jedrzejewski, MD, Align Surgical Associates Inc, San Francisco, California (data collection); Emery Potter, NP-PHC, Transition-related Surgery Program, Women's College Hospital, Toronto, Ontario, Canada (data collection); Edwin Wilkins, MD, Section of Plastic Surgery, University of Michigan, Ann Arbor (site supervision, data collection); Jennifer Hamill, MPH, Section of Plastic Surgery, University of Michigan, Ann Arbor (data collection); Sophia Silverman, MPH, Section of Plastic Surgery, University of Michigan, Ann Arbor (data collection); Amanda Higgins, MPH, Department of Surgery, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts (data collection); Navneet Pabla, BS, Patient-Reported Outcomes and Values, & Experience Center (PROVE), Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts (data collection); Amitai Miller, MD/MPP candidate, Harvard Medical School | Harvard Kennedy School, Boston, Massachusetts (data collection); Jason Kroening-Roche, MD, Primary Care Clinic, Oregon Health & Science University, Portland (data collection); Asha Jetmalani, MD, Transgender Health Program, Oregon Health & Science University, Portland (data collection); Daniel Dugi, MD, Urology Clinic, Oregon Health & Science University, Portland (data collection); Heather Onoday, RN, Dermatology Clinic Oregon Health & Science University, Portland (data collection); Juliana Hansen, MD, Division of Plastic and Reconstructive Surgery, Oregon Health & Science University, Portland (data collection); Sarah Skog, MD, Primary Care Clinic, Oregon Health & Science University, Portland (data collection); Mair Marsiglio, PhD, Transgender Health Program, Oregon Health & Science University, Portland (data collection); Emily Jacobsen, MD, Primary Care Clinic, Oregon Health & Science University, Portland (data collection); Dorian Scull, PA-C, Urology Clinic, Oregon Health & Science University, Portland (data collection); Amy Penkin, MSW, Transgender Health Program, Oregon Health & Science University, Portland (data collection); Mujde Özer, MD, BovenIJ ziekenhuis, Amsterdam, the Netherlands (data collection); Maeghan B. Ross, MSc, Amsterdam University Medical Center, Amsterdam, the Netherlands (data collection); Andrew Cohen, MD, Division of Urology, Johns Hopkins, Baltimore, Maryland (data collection); Errol Fields, MD, PhD, Center for Transgender and Gender Expansive Health, Johns Hopkins, Baltimore, Maryland (data collection); Helene F. Hedian, MD, Center for Transgender and Gender Expansive Health, Johns Hopkins, Baltimore, Maryland (data collection); Kate Thomas, PhD, Center for Transgender and Gender Expansive Health, Johns Hopkins, Baltimore, Maryland (data collection); Arthur Burnett, MD, Division of Urology, Johns Hopkins, Baltimore, Maryland (data collection); Dennis L. Tran, BS, GU Recon, San Francisco, California (data collection); and Sthefano Araya, MD, Temple University, Lewis Katz School of Medicine, Philadelphia, Pennsylvania (data collection). They were not compensated for their contributions.

REFERENCES

1. World Health Organization. Gender incongruence and transgender health in the ICD. Accessed June 12, 2024. https://www.who.int/standards/classifications/frequently-asked-questions/gender-incongruence-and-transgender-health-in-the-icd

2. Lavallee DC, Chenok KE, Love RM, et al. Incorporating patient-reported outcomes into health care to engage patients and enhance care. *Health Aff (Millwood)*. 2016;35(4):575-582. doi:10.1377/hlthaff.2015.1362

3. US Food and Drug Administration. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. December 2009. Accessed May 12, 2024. https://www.fda.gov/media/77832/download

4. Aiyegbusi OL, Hughes SE, Calvert MJ. The role of patient-reported outcomes (PROs) in the improvement of healthcare delivery and service. In: Kassianos AP, ed. *Handbook of Quality of Life in Cancer*. Springer; 2022:339-352.

5. Snyder CF, Aaronson NK, Choucair AK, et al. Implementing patient-reported outcomes assessment in clinical practice: a review of the options and considerations. *Qual Life Res*. 2012;21(8):1305-1314. doi:10.1007/s11136-011-0054-x

6. Fares CM, Williamson TJ, Theisen MK, et al. Low concordance of patient-reported outcomes with clinical and clinical trial documentation. *JCO Clin Cancer Inform*. 2018;2:1-12. doi:10.1200/CCI.18.00059

7. National Quality Forum. National Voluntary Consensus Standards for Patient Outcomes 2009. July 2011. Accessed August 24, 2024. https://www.qualityforum.org/Publications/2011/07/National_Voluntary_Consensus_Standards_for_Patient_Outcomes_2009.aspx

8. Deutsch ASL, Gage B, Kelleher C, Garfinkel D. *Patient-Reported Outcomes in Performance Measurement*. RTI Press; 2012.

9. US Food and Drug Administration. Framework for FDA's real-world evidence program. December 2018. Accessed May 12, 2024. https://www.fda.gov/media/120060/download

10. Kluetz PG, O'Connor DJ, Soltys K. Incorporating the patient experience into regulatory decision making in the USA, Europe, and Canada. *Lancet Oncol*. 2018;19(5):e267-e274. doi:10.1016/S1470-2045(18)30097-4

11. Centers for Medicare & Medicaid Services. CMS national quality strategy. Updated February 28, 2025. Accessed March 3, 2025. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/CMS-Quality-Strategy

12. Arlett P, Kjaer J, Broich K, Cooke E. Real-world evidence in EU medicines regulation: enabling use and establishing value. *Clin Pharmacol Ther*. 2022;111(1):21-23. doi:10.1002/cpt.2479

13. UK Government's Medicines and Healthcare Products Regulatory Agenda. MHRA guideline on randomised controlled trials using real-world data to support regulatory decisions. December 16, 2021. Accessed March 4, 2025. https://www.gov.uk/government/publications/mhra-guidance-on-the-use-of-real-world-data-in-clinical-studies-to-support-regulatory-decisions/mhra-guideline-on-randomised-controlled-trials-using-real-world-data-to-support-regulatory-decisions

14. UK Government's Medicines and Healthcare products Regulatory Agenda. MHRA guidance on the use of real-world data in clinical studies to support regulatory decisions. December 16, 2021. Accessed March 4, 2025. https://www.gov.uk/government/publications/mhra-guidance-on-the-use-of-real-world-data-in-clinical-studies-to-support-regulatory-decisions/mhra-guidance-on-the-use-of-real-world-data-in-clinical-studies-to-support-regulatory-decisions

15. Baker KE, Wilson LM, Sharma R, Dukhanin V, McArthur K, Robinson KA. Hormone therapy, mental health, and quality of life among transgender people: a systematic review. *J Endocr Soc*. 2021;5(4):bvab011. doi:10.1210/jendso/bvab011

16. Bränström R, Pachankis JE. Reduction in mental health treatment utilization among transgender individuals after gender-affirming surgeries: a total population study. *Am J Psychiatry*. 2020;177(8):727-734. doi:10.1176/appi.ajp.2019.19010080

17. White Hughto JM, Reisner SL. A systematic review of the effects of hormone therapy on psychological functioning and quality of life in transgender individuals. *Transgend Health*. 2016;1(1):21-31. doi:10.1089/trgh.2015.0008

18. Witcomb GL, Bouman WP, Claes L, Brewin N, Crawford JR, Arcelus J. Levels of depression in transgender people and its predictors: results of a large matched control study with transgender people accessing clinical services. *J Affect Disord*. 2018;235:308-315. doi:10.1016/j.jad.2018.02.051

19. Manrique OJ, Adabi K, Martinez-Jorge J, Ciudad P, Nicoli F, Kiranantawat K. Complications and patient-reported outcomes in male-to-female vaginoplasty—where we are today: a systematic review and meta-analysis. *Ann Plast Surg*. 2018;80(6):684-691. doi:10.1097/SAP.0000000000001393

20. 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(1):e52-e66. doi:10.1097/SLA.0000000000004728

21. 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(1):e67-e74. doi:10.1097/SLA.0000000000004717

22. Tolstrup A, Zetner D, Rosenberg J. Outcome measures in gender-confirming chest surgery: a systematic scoping review. *Aesthetic Plast Surg*. 2020;44(1):219-228. doi:10.1007/s00266-019-01523-1

23. Caprini RM, Oberoi MK, Dejam D, et al. Effect of gender-affirming facial feminization surgery on psychosocial outcomes. *Ann Surg*. 2023;277(5):e1184-e1190. doi:10.1097/SLA.0000000000005472

24. 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(6):1499-1509. doi:10.1097/PRS.0000000000006837

25. Boskey ER, Jolly D, Kant JD, Ganor O. Prospective evaluation of psychosocial changes after chest reconstruction in transmasculine and non-binary youth. *J Adolesc Health*. 2023;73(3):503-509. doi:10.1016/j.jadohealth.2023.04.029

26. Clennon EK, Martin LH, Fadich SK, et al. Community engagement and patient-centered implementation of patient-reported outcome measures (PROMs) in gender affirming surgery: a systematic review. *Curr Sex Health Rep*. 2022;14(1):17-29. doi:10.1007/s11930-021-00323-6

27. Dy GW, Nolan IT, Hotaling J, Myers JB. Patient reported outcome measures and quality of life assessment in genital gender confirming surgery. *Transl Androl Urol*. 2019;8(3):228-240. doi:10.21037/tau.2019.05.04

28. 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(4):1026-1039. doi:10.1097/PRS.0000000000004254

29. Barone M, Cogliandro A, Persichetti P. Patient-reported outcome measures used in gender confirmation surgery: a systematic review. *Plast Reconstr Surg*. 2018;142(6):985e-986e. doi:10.1097/PRS.0000000000005035

30. Terwee CB, Prinsen C, Chiarotto A, et al. COSMIN Methodology for Assessing the Content Validity of PROMs—User Manual. February 2018. Accessed March 4, 2025. https://www.cosmin.nl/wp-content/uploads/COSMIN-methodology-for-content-validity-user-manual-v1.pdf

31. Gagnier JJ, Lai J, Mokkink LB, Terwee CB. COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures. *Qual Life Res*. 2021;30(8):2197-2218. doi:10.1007/s11136-021-02822-4

32. 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(8):1889-1905. doi:10.1007/s11136-012-0344-y

33. Mokkink LB, Prinsen C, Patrick DL, et al. COSMIN study design checklist for patient-reported outcome measurement instruments. July 2019. Accessed March 4, 2025. https://gut.bmj.com/content/gutjnl/70/1/139/DC1/embed/inline-supplementary-material-1.pdf

34. Kaur MN, Morrison SD, Kennedy SL, 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(1):134. doi:10.1186/s41687-024-00785-x

35. 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(10):e025435. doi:10.1136/bmjopen-2018-025435

36. Eremenco S, Pease S, Mann S, Berry P; PRO Consortium's Process Subcommittee. Patient-Reported Outcome (PRO) Consortium translation process: consensus development of updated best practices. *J Patient Rep Outcomes*. 2017;2(1):12. doi:10.1186/s41687-018-0037-6

37. Wild D, Grove A, Martin M, et al; ISPOR Task Force for Translation and Cultural Adaptation. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR task force for translation and cultural adaptation. *Value Health*. 2005;8(2):94-104. doi:10.1111/j.1524-4733.2005.04054.x

38. Knottnerus A, Tugwell P. STROBE—a checklist to Strengthen the Reporting of Observational Studies in Epidemiology. *J Clin Epidemiol*. 2008;61(4):323. doi:10.1016/j.jclinepi.2007.11.006

39. Trans PULSE. Accessed March 4, 2025. http://transpulseproject.ca/

40. Rasch G. A mathematical theory of objectivity and its consequences for model construction. European Meeting on Statistics, Econometrics and Management Sciences, Amsterdam; September 2-7, 1968. Accessed March 4, 2025. https://rasch.org/memo1968.pdf

41. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1147-1157. doi:10.1007/s11136-018-1798-3

42. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155-163. doi:10.1016/j.jcm.2016.02.012

43. De Vet HC, Terwee CB, Mokkink LB, Knol DL. *Measurement in Medicine: A Practical Guide*. Cambridge University Press; 2011. doi:10.1017/CBO9780511996214

44. Grainger C. Rasch analysis is important to understand and use for measurement. Accessed March 4, 2025. https://www.rasch.org/rmt/rmt213d.htm

45. Cano SJ, Hobart JC. The problem with health measurement. *Patient Prefer Adherence*. 2011;5:279-290. doi:10.2147/PPA.S14399

46. Hughes S, Aiyegbusi OL, Lasserson D, Collis P, Glasby J, Calvert M. Patient-reported outcome measurement: a bridge between health and social care? *J R Soc Med*. 2021;114(8):381-388. doi:10.1177/01410768211014048

47. Al Sayah F, Lahtinen M, Bonsel GJ, Ohinmaa A, Johnson JA. A multi-level approach for the use of routinely collected patient-reported outcome measures (PROMs) data in healthcare systems. *J Patient Rep Outcomes*. 2021;5(suppl 2):98. doi:10.1186/s41687-021-00375-1

48. Gonçalves Bradley DC, Gibbons C, Ricci-Cabello I, et al. Routine provision of information on patient-reported outcome measures to healthcare providers and patients in clinical practice. *Cochrane Database Syst Rev*. Published online April 29, 2015. doi:10.1002/14651858.CD011589

49. Bonsel JM, Itiola AJ, Huberts AS, Bonsel GJ, Penton H. The use of patient-reported outcome measures to improve patient-related outcomes: a systematic review. *Health Qual Life Outcomes*. 2024;22(1):101. doi:10.1186/s12955-024-02312-4

50. Crow R, Gage H, Hampson S, et al. The measurement of satisfaction with healthcare: implications for practice from a systematic review of the literature. *Health Technol Assess*. 2002;6(32):1-244. doi:10.3310/hta6320

51. Ware JE Jr, Hays RD. Methods for measuring patient satisfaction with specific medical encounters. *Med Care*. 1988;26(4):393-402. doi:10.1097/00005650-198804000-00008

52. Voutilainen A, Pitkäaho T, Kvist T, Vehviläinen-Julkunen K. How to ask about patient satisfaction? the visual analogue scale is less vulnerable to confounding factors and ceiling effect than a symmetric Likert scale. *J Adv Nurs*. 2016;72(4):946-957. doi:10.1111/jan.12875

53. Bjertnaes O, Iversen HH, Garratt AM. The Universal Patient Centeredness Questionnaire: scaling approaches to reduce positive skew. *Patient Prefer Adherence*. 2016;10:2255-2260. doi:10.2147/PPA.S116424

54. Q Portfolio. Accessed March 4, 2025. https://qportfolio.org/

SUPPLEMENT 1.

eTable 1. Field Test Sites
eTable 2. RMT Criteria and Statistical Tests
eTable 3. Test-Retest Reliability Results
eTable 4. Construct Validity
eTable 5. Distribution of Urinary Function Checklist
eTable 6. Distribution of Surgery, Adverse Effect Checklist
eTable 7. Mean Scores and Key Demographic Characteristics for Scales

SUPPLEMENT 2.

Data Sharing Statement

Supplemental Online Content

Kaur MN, Rae C, Morrison SD, et al. Development and assessment of a patient-reported outcome instrument for gender-affirming care. *JAMA Netw Open.* 2025;8(4):e254708. doi:10.1001/jamanetworkopen.2025.4708

This supplemental material has been provided by the authors to give readers additional information about their work.

eTable 1. Field Test Sites

	Organization name/location	Survey distribution method	REDCap survey hosted at McMaster University
Crowdsourcing / existing research databases	Prolific	Participants accessed survey through Prolific platform	Yes
	Transpulse Survey Canada, Western University, London, Ontario	Emailed survey link to database participants	Yes
Social Media Recruitment Only	Flinders University, Adelaide, South Wales, Australia	Shared survey link through X (Twitter), Facebook, Reddit, Discord platforms for LGBTQ2+ community support groups	Yes
	Odense University Hospital, Research Unit for Plastic Surgery, Odense, Denmark	Shared survey link at Copenhagen PRIDE event Shared survey link through Facebook groups for LGBTQ2+ community support groups	Danish survey – No English survey – Yes
	East of England Gender Service, Cambridge, United Kingdom	Shared survey link through LGBTQ2+ community support groups (e.g., OUTpatients)	Yes
Clinical Sites			
Canada	GrS Montreal, Montreal, Quebec	Emailed survey link to retrospective patient list	Yes
	McLean Clinic, Mississauga, Ontario	Emailed survey link to retrospective patient list	Yes
	Women's College Hospital, Trans-related Surgery Center, Toronto, Ontario	Emailed survey link to retrospective patient list	Yes
United States	Align Surgical Associates, San Francisco, California	Emailed survey link to retrospective patient list	Yes
	Boston Children's Hospital, Center for Gender Surgery, Boston, Massachusetts	Emailed survey link to retrospective patient list	No
	Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts	Emailed survey link to retrospective patient list	No
	Cedars-Sinai Medical Center, Department of Surgery, Los Angeles, California	Emailed survey link to retrospective patient list	Yes
	Crane Center for Transgender Surgery, Austin, Texas	Emailed survey link to retrospective patient list	Yes
	G.U. Recon Clinic, San Francisco, California	Emailed survey link to retrospective patient list	Yes
	Johns Hopkins, Baltimore Maryland	Emailed survey link to retrospective patient list	Yes
	MedStar Georgetown University Hospital, Washington D.C.	Emailed survey link to retrospective patient list In-person recruitment in clinic	Yes
	New York University Langone Health, New York, New York	Emailed survey link to retrospective patient list	No
	Oregon Health and Sciences University, Portland, Oregon	Emailed survey link to retrospective patient list	Yes

	Rebirth Health Center, Salt Lake City, Utah	Advertising flyers in clinic Social media – posted survey link on clinic's Facebook page	Yes
	Rush University Medical Center, Rush Gender Affirmation Surgery, Chicago, Illinois	Advertising flyers in clinic	Yes
	Temple University, Lewis Katz School of Medicine, Philadelphia, Pennsylvania	Emailed survey link to retrospective patient list	Yes
	University of Colorado, UCHealth Plastic and Reconstructive Surgery Clinic – Anschutz Medical Campus, Aurora, Colorado	Emailed survey link to retrospective patients In-person recruitment in clinic or by telephone with current patients	No
	University of Michigan, Department of Plastic Surgery, Ann Arbor, Michigan	Emailed survey link to retrospective patient list	No
	Yale University, Plastic & Reconstructive Surgery, New Haven, Connecticut	In-person recruitment in clinic	No
The Netherlands	Amsterdam University Medical Center, Center for Expertise on Gender Dysphoria, Amsterdam	Emailed survey link to retrospective patient list In-person recruitment in clinic	No
Spain	FacialTeam Group, HC Marbella International Hospital, Malaga	Emailed survey link to retrospective patient list Social media – posted survey link on clinic's X (Twitter), Instagram, LinkedIn and Facebook pages Survey link shared during Live interview with GENDER-Q Study Investigators on FacialTeam members-only Youtube Channel	Yes

eTable 2. RMT Criteria and Statistical Tests

Response threshold order – evaluates if the response categories are ordered such that as the latent trait (ability) increases, the probability of endorsing a response category aligned with more of the latent trait also increases in a predictable manner, i.e., more individuals should endorse a lower level of the latent trait and fewer should endorse a higher level of the latent trait. The number of thresholds is equal to the number of response options minus one. Disordered thresholds may indicate that there are too many response options, or that the labels for response options are confusing.
Item fit – evaluates the extent to which the observed data fit the expectations of the Rasch model. When data fit the model, the items should have a hierarchical order such that items that indicate lower levels of latent trait are at the lower end of the Rasch "ruler" and items that indicate higher levels of the latent trait are at the higher end. Item fit is indicated by non-significant Chi-square after Bonferroni adjustment and fit residuals that are ideally within the range -2.5 and +2.5. Item fit is also examined graphically using item characteristic curves that show the probability of a correct response across different levels of the latent trait. The goal is to have well-fitting items with smooth, monotonically increasing item characteristic curves.
Local dependency – indicates the extent to which the response to an item influences the response to another item in an item set. The inter-relatedness of items, if detected, impacts the probabilistic structure of the Rasch model and inflates the other psychometric properties of the scale (i.e., reliability and validity). Items with pair-wise residual correlations higher than 0.3 indicate local dependence. Locally dependent items are evaluated in a subtest to determine their impact on scale's reliability.
Targeting – evaluates the spread of person locations and item locations. Person locations for a well targeted scale are centred at zero and have a standard deviation of 1. This is inspected graphically with person-item threshold plots with the goal to have limited clustering of items and gaps on the scale. A scale that is well targeted has more coverage and has the mean person location close to the center of the items. The proportion of a sample that scores on the scale's range of measurement can also be determined.

Differential item functioning (DIF) – evaluates if the item difficulty hierarchy is consistent across subgroups of people being measured. DIF is assessed with a significant F-test from a two-way analysis of variance and graphically using item characteristic curves. For GENDER-Q, DIF was examined for the following characteristics: age (18-24, 25-29, 30-39, 40-49, 50 years) and the goal of gender-affirming care (masculinizing or feminizing). DIF was computed for scales after selecting a random sample of equal sized subgroups. DIF was performed as long as the subgroups included at least 50 participants, and the overall sample size was at least 300. Items that evidenced significant was split on the sample characteristic. Person correlations between the original and the new person locations were computed to examine the impact of DIF on scoring.
Reliability - indicates how well a scale can distinguish between different levels of person's ability. This form of reliability is assessed with the person separation index (PSI), where higher values indicate better discrimination. A scale with higher PSI values facilitates measurement of change. Values >0.9 are considered suitable for measuring within-person change and values >0.7 are suitable for detecting group differences. Cronbach alpha is used to examine internal reliability. For test-retest reliability, intraclass correlation coefficients (ICC) with a two-way random effects model are used to evaluate the consistency of responses 7–14 days after the initial scale completion, barring changes in status for the construct measured by the scale.
Hypothesis-based construct validity - refers to the extent to which the scale accurately measures the construct it purports to measure. Parametric or nonparametric tests were used depending on the distribution of the data. Rasch transformed scale scores (0-worse, 100-best) were used. It is of major importance that the hypotheses are defined in advance when assessing construct validity to enable one to draw unbiased conclusions after data collection.

eTable 3. Test-Retest Reliability Results

Scale	N	Valid n	Single ICC	95% Confidence interval		Average ICC	95% Confidence interval		Means						SEM	SDC individual	SDC group
				LB	UB		LB	UB	T1	T1 SD	T2	T2 SD	Mean diff	Mean diff SD	Sd_{pooled}* √(1 – ICC)	1.96*√2 *SEM	SDC_{ind} /√n
HEALTH-RELATED QUALITY OF LIFE																	
Body Image	125	106	0.77	0.68	0.84	0.87	0.81	0.91	43	21	44	21	1.8	14.2	7.6	21.0	2.0
Gender Dysphoria	125	87	0.84	0.76	0.89	0.91	0.86	0.94	53	17	56	18	2.6	9.6	5.2	14.3	1.5
Social Acceptance	125	93	0.79	0.70	0.86	0.88	0.82	0.92	64	12	64	13	0.0	8.0	4.2	11.7	1.2
Psychological Distress	125	87	0.74	0.63	0.82	0.85	0.77	0.90	55	19	56	20	0.9	14.2	7.6	21.0	2.3
Psychological Well-Being	125	88	0.73	0.62	0.82	0.85	0.77	0.90	50	17	51	17	1.5	12.4	6.7	18.6	2.0
Treatment Outcome	144	123	0.81	0.73	0.86	0.89	0.85	0.93	71	21	68	22	-3.2	13.0	7.0	19.3	1.7
SEXUAL																	
Sexual Well-Being	125	106	0.79	0.71	0.85	0.88	0.83	0.92	55	17	55	14	-0.2	10.2	5.4	15.0	1.5
Orgasm	125	106	0.85	0.78	0.90	0.92	0.87	0.95	61	17	64	18	2.8	9.1	5.0	13.8	1.3
GENDER PRACTICES																	
Binding Well-Being	111	107	0.74	0.60	0.83	0.85	0.75	0.91	66	17	61	18	-5.1	12.0	6.8	18.9	1.8
Binding Adverse - Body	110	100	0.77	0.67	0.84	0.87	0.80	0.91	68	16	65	16	-2.8	10.3	5.7	15.8	1.6
Binding Adverse - Skin	110	100	0.71	0.60	0.80	0.83	0.75	0.89	73	20	72	18	-1.9	14.5	7.9	21.8	2.2
Tucking Adverse	36	29	0.76	0.56	0.88	0.87	0.72	0.94	75	19	73	17	-1.8	12.5	6.6	18.3	3.4
VOICE																	
Sound	125	111	0.80	0.71	0.86	0.89	0.83	0.92	48	14	50	18	2.1	10.2	5.4	15.0	1.4
Distress	125	111	0.79	0.71	0.85	0.89	0.83	0.92	57	19	59	21	2.0	13.1	6.9	19.0	1.8
HAIR																	
Face- Feminization	110	28	0.91	0.82	0.96	0.95	0.90	0.98	42	16	43	15	0.8	6.6	3.4	9.3	1.8
Face- Masculinization	108	47	0.76	0.61	0.86	0.86	0.76	0.92	50	18	51	21	1.4	13.6	7.2	19.9	2.9
Head	124	105	0.85	0.77	0.90	0.92	0.87	0.94	62	20	59	19	-3.0	10.6	5.7	15.8	1.5
FACE & NECK																	
Face Overall	124	105	0.82	0.75	0.88	0.90	0.86	0.93	47	14	48	16	1.1	9.1	4.7	13.1	1.3
Facial Features	124	105	0.79	0.71	0.85	0.88	0.83	0.92	53	13	54	13	1.3	8.5	4.4	12.3	1.2
Upper Face	124	105	0.77	0.68	0.84	0.87	0.81	0.91	57	18	56	19	-1.0	12.3	6.6	18.3	1.8
Eyebrows	124	105	0.83	0.76	0.88	0.91	0.86	0.94	61	20	63	18	1.9	11.1	5.9	16.2	1.6

Cheeks	124	105	0.82	0.74	0.87	0.90	0.85	0.93	54	20	54	21	0.5	12.3	6.4	17.8	1.7
Nose	124	105	0.83	0.76	0.88	0.91	0.87	0.94	57	21	58	21	0.3	12.1	6.3	17.4	1.7
Nostrils	124	105	0.74	0.64	0.82	0.85	0.78	0.90	63	20	63	21	-0.1	14.9	7.9	21.8	2.1
Lips	124	105	0.79	0.71	0.86	0.89	0.83	0.92	59	19	61	20	1.7	12.3	6.6	18.2	1.8
Chin	124	105	0.75	0.66	0.83	0.86	0.79	0.91	53	21	52	22	-0.9	15.2	8.1	22.4	2.2
Jawline	124	105	0.85	0.78	0.89	0.92	0.88	0.94	47	22	49	22	1.7	12.2	6.3	17.6	1.7
Adam's Apple	97	83	0.90	0.85	0.93	0.95	0.92	0.97	64	27	64	26	-0.6	12.1	6.1	17.0	1.9
BODY																	
Body	125	106	0.81	0.73	0.87	0.89	0.84	0.93	38	17	40	18	2.3	10.5	5.7	15.7	1.5
Buttocks	124	105	0.88	0.82	0.92	0.94	0.90	0.96	46	22	50	23	3.1	10.8	5.7	15.8	1.5
Waist	124	105	0.75	0.65	0.82	0.86	0.79	0.90	38	23	42	21	3.1	15.4	8.4	23.2	2.3
BREAST																	
Breast	133	107	0.86	0.80	0.90	0.92	0.89	0.95	50	19	52	19	1.6	9.9	5.2	14.4	1.4
Nipples & Areolas	132	107	0.85	0.79	0.90	0.92	0.88	0.94	61	20	61	19	-0.3	10.8	5.5	15.4	1.5
GENITAL FEMINIZATION																	
Vagina	39	27	0.89	0.77	0.95	0.94	0.87	0.97	57	18	58	22	0.6	9.6	4.9	13.5	2.6
Labia	38	26	0.89	0.77	0.95	0.94	0.87	0.97	56	15	58	15	1.8	7.0	3.7	10.3	2.0
Clitoris	37	25	0.87	0.74	0.94	0.93	0.85	0.97	63	21	65	25	1.4	11.9	6.0	16.7	3.3
Dilation	33	21	0.69	0.37	0.86	0.82	0.54	0.93	63	16	62	14	-1.0	12.3	6.5	18.1	4.0
CHEST																	
Chest	226	208	0.93	0.91	0.95	0.96	0.95	0.97	49	31	49	27	0.2	11.0	5.6	15.6	1.1
Nipples & Areolas	226	209	0.88	0.84	0.90	0.93	0.91	0.95	53	21	55	21	1.3	10.3	5.4	15.0	1.0
Scars	112	97	0.90	0.86	0.94	0.95	0.93	0.97	80	18	79	17	-0.9	7.7	3.9	10.8	1.1
GENTIAL MASCULINIZATION																	
Penis	42	30	0.90	0.79	0.95	0.95	0.89	0.97	62	15	63	16	0.6	7.1	3.6	10.0	1.8
Penis Sensation	41	34	0.96	0.93	0.98	0.98	0.96	0.99	60	28	63	26	2.1	7.2	3.7	10.3	1.8
Glans	26	18	0.89	0.72	0.96	0.94	0.83	0.98	72	21	76	20	4.0	9.2	5.0	13.8	3.3
Scrotum	34	25	0.91	0.81	0.96	0.95	0.90	0.98	66	19	66	16	0.6	7.5	3.8	10.6	2.1
Perineum	23	19	0.82	0.59	0.93	0.90	0.75	0.96	69	24	72	25	3.4	14.9	7.8	21.7	5.0
Donor Site - Forearm or Thigh	28	21	0.48	0.08	0.75	0.65	0.14	0.86	66	25	71	21	5.0	24.1	13.8	38.3	8.4
Donor Site - Adverse Effects	27	20	0.57	0.18	0.81	0.73	0.30	0.89	80	16	80	14	0.8	14.0	7.8	21.7	4.9
Testicular Implants	12	11	0.88	0.62	0.97	0.94	0.76	0.98	62	16	62	19	0.2	8.8	4.4	12.2	3.7
Erectile Device	10	6	0.90	0.50	0.99	0.95	0.67	0.99	60	15	57	16	-2.7	7.3	3.7	10.1	4.1

EXPERIENCE CARE																	
Health Professional	152	80	0.72	0.59	0.81	0.84	0.75	0.90	78	22	76	22	-2.3	16.5	8.9	24.6	2.8
Clinic	105	58	0.78	0.65	0.86	0.87	0.79	0.93	79	21	78	19	-1.2	13.4	7.1	19.7	2.6
Surgery - Return to Activity	11	9	0.57	-0.02	0.88	0.72	-0.04	0.94	88	16	96	10	7.9	11.4	7.1	19.6	6.5

ICC, Intraclass correlation co-efficient; UB, upper bound; LB, lower bound; T1, time 1; T2, time 2 (7-14 days after T1); SEM, standard error of measurement; SDC, smallest detectable change; SD, standard deviation

eTable 4. Construct Validity

eTable 4a. RESULTS FOR HYPOTHESIS-BASED CONSTRUCT VALIDITY TESTING OF THE GENDER-Q SCALES

Scale	Response	N	Mean	Standard deviation	Standard error	95% CI		p-value
						LB	UB	
In general, how would you rate your satisfaction with your social activities and relationships?								
Gender Dysphoria	Poor	628	49	19	1	48	51	<0.001
	Fair	1081	58	18	1	57	59	
	Good	1394	63	19	1	62	64	
	Very Good	967	69	19	1	68	70	
	Excellent	384	78	21	1	76	80	
Social Acceptance	Poor	647	60	17	1	59	61	<0.001
	Fair	1107	68	15	0	67	69	
	Good	1428	74	16	0	74	75	
	Very Good	988	80	15	0	79	81	
	Excellent	397	88	13	1	87	89	
In general, how would you rate your mental health, including your mood and your ability to think?								
Psychological Distress	Poor	644	48	21	1	46	50	<0.001
	Fair	1211	59	19	1	58	60	
	Good	1291	67	18	1	66	68	
	Very Good	905	74	17	1	72	75	
	Excellent	381	83	17	1	81	85	
Psychological Well-Being	Poor	661	45	18	1	43	46	<0.001
	Fair	1233	55	16	0	54	56	
	Good	1324	64	16	0	63	65	
	Very Good	927	72	16	1	71	73	
	Excellent	386	85	16	1	83	86	
How satisfied are you with your sex life?								
Sexual Well-Being	Not at all satisfied	416	45	14	1	43	46	<0.001
	A little satisfied	705	51	12	0	50	52	

	Somewhat satisfied	1293	58	12	0	57	58	
	Very satisfied	952	67	15	0	66	68	
	Extremely satisfied	462	77	17	1	75	78	
When you engage in sexual activity (e.g., masturbation, partnered sex), are you able to have an orgasm?								
Orgasm	Rarely	224	37	16	1	35	39	<0.001
	Sometimes	666	50	13	1	49	51	
	Often	1432	61	12	0	61	62	
	Always	1381	75	17	0	74	76	
Your donor area overall?								
Donor Site - Forearm or Thigh	Not at all bothered	74	88	13	1	85	91	<0.001
	A little bothered	91	66	10	1	64	68	
	Somewhat bothered	42	53	7	1	51	56	
	Very bothered	27	44	10	2	40	48	
	Extremely bothered	17	21	23	6	9	33	
Your donor area scar overall?								
Donor Site - Adverse Effects	Not at all concerned	149	88	11	1	86	90	<0.001
	A little concerned	67	67	10	1	64	69	
	Somewhat concerned	22	62	17	4	55	69	
	Very concerned	7	63	18	7	46	80	
	Extremely concerned	6	33	20	8	13	54	
How much feeling do you have in your penis?								
Penis Sensation	I have no feeling	30	11	14	3	6	16	<0.001
	I have a little feeling	70	32	11	1	30	35	
	I have some feeling	97	45	10	1	43	47	
	I have a lot of feeling	97	62	13	1	59	64	
	I have complete feeling	87	90	15	2	86	93	
Your chest scars overall?								
Chest - Scars	Not at all bothered	1080	91	11	0	90	92	<0.001
	A little bit bothered	668	68	10	0	68	69	
	Quite a bit bothered	118	56	11	1	54	58	

	Very much bothered	54	36	21	3	30	41	
How your Adam's apple looks overall?								
Adam's Apple	Not at all bothered	1031	98	5	0	98	99	<0.001
	A little bothered	456	72	11	0	71	73	
	Somewhat bothered	238	53	10	1	52	55	
	Very bothered	131	41	14	1	39	43	
	Extremely bothered	94	19	18	2	15	22	
Overall, I am completely satisfied with the care I received from my health professional.								
Health Professional	Strongly disagree	64	29	17	2	25	34	p<0.001
	Mostly disagree	55	44	9	1	41	46	
	Slightly disagree	71	54	8	1	52	56	
	Slightly agree	140	60	11	1	59	62	
	Mostly agree	584	75	13	1	74	76	
	Strongly agree	2086	97	7	0	96	97	
Overall, I am completely satisfied with the clinic.								
Clinic	Strongly disagree	33	34	23	4	25	42	p<0.001
	Mostly disagree	45	49	14	2	45	53	
	Slightly disagree	65	55	13	2	52	58	
	Slightly agree	164	63	13	1	61	65	
	Mostly agree	510	76	14	1	75	78	
	Strongly agree	1506	97	7	0	97	97	
Overall, I am completely satisfied with the information I received about my gender-affirming surgery.								
Surgery - Information	Not at all satisfied	14	36	12	3	29	44	p<0.001
	A little satisfied	32	47	7	1	44	49	
	Somewhat satisfied	81	55	9	1	53	57	
	Very satisfied	165	68	11	1	66	70	
	Extremely satisfied	236	94	11	1	93	95	
Overall, I am completely satisfied with my gender-affirming treatment.								
Treatment Outcome	Strongly disagree	113	26	19	2	23	30	p<0.001
	Mostly disagree	88	37	16	2	34	40	

	Slightly disagree	179	48	13	1	46	50	
	Slightly agree	265	54	11	1	52	55	
	Mostly agree	917	68	12	0	67	68	
	Strongly agree	1893	90	12	0	89	90	

UB, upper bound; LB, lower bound; CI, confidence interval

eTable 4b. RESULTS FOR HYPOTHESIS-BASED CONSTRUCT VALIDITY TESTING OF THE GENDER-Q SCALES MEASURING SATISFACTION WITH APPEARANCE

Scale All tests significant p<0.001	Overall Question	Extremely dissatisfied					Very dissatisfied					Somewhat dissatisfied					Somewhat satisfied					Very satisfied					Extremely satisfied				
		N	Mean	SD	95% CI		N	Mean	SD	95% CI		N	Mean	SD	95% CI		N	Mean	SD	95% CI		N	Mean	SD	95% CI		N	Mean	SD	95% CI	
					LB	UB				LB	UB				LB	UB				LB	UB				LB	UB				LB	UB
Voice Sound	How your voice sounds overall?	498	24	13	23	25	695	37	5	37	37	972	44	5	44	45	1320	51	5	51	51	1237	62	6	62	62	661	83	14	82	84
Voice Distress	How your voice sounds overall?	494	35	17	34	37	684	45	12	45	46	966	55	11	54	55	1311	64	14	63	65	1227	79	14	78	79	648	88	14	87	90
Face Overall	How your face looks overall?	264	22	12	20	23	408	35	5	35	36	811	43	5	42	43	1500	52	6	52	53	1289	66	7	65	66	598	91	11	90	92
Facial Features	How your face looks overall?	262	36	14	35	38	405	44	7	43	44	806	49	8	48	49	1486	55	9	55	56	1271	66	11	65	67	597	89	13	88	90
Upper Face	How your upper face looks overall?	199	25	15	23	27	321	38	9	37	39	612	47	9	47	48	1298	57	8	57	58	1369	70	8	70	71	951	94	10	93	94
Eyebrows	How your eyebrows look overall?	76	19	16	15	22	140	38	8	37	39	417	45	8	44	46	982	54	7	53	54	1456	71	7	70	71	1158	97	7	96	97
Cheeks	How your cheeks look overall?	114	9	10	7	11	117	27	6	26	29	276	40	6	39	40	433	54	6	53	54	252	72	7	71	73	115	94	10	92	96
Nose	How your nose looks overall?	213	15	13	14	17	266	33	7	32	34	379	43	6	42	43	467	53	6	53	54	316	69	7	68	69	173	94	9	93	95
Nostrils	How your nostrils look overall?	120	12	16	9	15	124	30	6	29	31	279	40	6	40	41	562	54	6	53	54	439	73	7	72	74	229	98	6	97	99
Lips	How your lips look overall?	114	14	13	12	17	170	34	6	33	34	325	43	6	42	44	437	53	5	53	54	275	68	7	68	69	162	90	11	89	92
Chin	How your chin looks overall?	228	9	11	8	11	298	29	6	28	30	455	41	5	41	42	514	54	6	54	55	266	72	7	71	73	136	95	9	94	97
Jaw	How your jawline looks overall?	318	9	11	7	10	382	29	7	28	30	640	41	6	41	42	514	55	6	55	56	249	71	7	70	72	105	93	11	91	95
Facial hair feminization	How your facial hair looks overall?	411	21	12	20	22	223	31	7	31	32	228	39	6	38	39	189	45	6	44	46	142	53	7	52	54	96	64	8	62	66
Facial hair masculinization	How your facial hair looks overall?	102	26	13	23	28	182	37	6	36	38	288	45	5	44	45	558	53	6	53	53	507	65	7	64	65	391	86	13	85	88
Head hair	How your hair looks overall?	224	21	16	19	23	241	39	7	38	39	434	47	6	46	47	906	53	5	53	54	1393	64	7	64	65	1387	90	13	89	90
Body	How your body looks overall?	521	19	13	18	20	687	34	7	33	34	1016	43	6	43	44	1377	52	6	52	53	987	65	8	65	66	362	88	11	87	89
Buttocks	How your buttocks look overall?	288	13	12	11	14	272	32	6	31	32	383	41	5	41	42	423	53	6	52	53	170	68	8	67	70	77	90	12	87	93
Waist	How your waist looks overall?	493	9	11	8	10	513	29	7	29	30	651	42	6	41	42	493	54	7	53	54	193	69	8	68	70	58	91	12	88	94
Breast	Your breasts overall?	138	28	16	26	31	201	43	6	42	44	305	48	5	47	49	526	54	6	53	54	605	63	6	62	63	343	80	13	78	81
Breast Nipples & Areolas	Your nipples and areolas overall?	78	26	19	22	30	112	44	7	43	45	229	48	8	47	50	531	55	7	54	55	637	66	8	66	67	478	91	12	90	92
Vagina	Your vagina overall?	37	15	16	10	20	39	31	10	27	34	93	33	14	31	36	209	43	9	42	44	458	55	10	54	56	386	75	17	73	77
Labia	Your labia overall?	36	23	16	17	28	43	33	10	30	36	93	44	6	43	45	283	50	6	49	51	391	60	9	59	61	294	82	17	80	84
Clitoris	Your clitoris overall?	26	17	17	10	24	49	34	15	29	38	83	43	12	40	45	180	49	11	48	51	378	61	13	60	62	393	83	18	81	85
Chest	How your chest looks overall?	399	16	13	15	18	195	34	10	33	36	142	46	9	45	48	320	59	10	57	60	816	73	12	72	74	975	92	11	91	92
Chest Nipples & Areolas	Your nipples and areolas overall?	245	27	17	25	29	201	41	7	40	42	335	47	8	46	48	553	54	8	53	54	651	66	9	65	66	540	90	12	89	91
Penis	Your penis overall?	12	20	16	10	30	15	35	11	29	41	32	44	7	42	47	73	54	8	52	56	141	63	8	61	64	113	79	14	77	82
Glans	Your glans overall?	14	26	18	16	37	10	34	8	28	40	12	47	7	42	52	48	54	10	51	57	73	63	10	61	66	66	84	15	80	87
Scrotum	Your scrotum overall?	11	20	15	10	30	16	38	7	34	42	27	43	6	41	46	74	51	5	49	52	95	62	8	60	64	83	81	15	77	84
												Dissatisfied					Somewhat satisfied					Very satisfied					Extremely satisfied				

Testicular Implants	Your testicular implants overall?	*Dissatisfied categories merged due to sample size	11	23	17	12	34	20	42	7	39	45	23	53	9	49	57	39	77	20	70	83
Perineum	Your perineum overall?		14	23	15	14	32	32	36	11	32	40	56	57	9	55	59	70	88	16	85	92
Erectile Device	Your erectile device overall?		17	36	10	31	41	19	43	9	38	47	24	56	8	53	59	17	73	11	68	79

eTable 5. Distribution of Urinary Function Checklist*

	Masculine appearance										Feminine appearance									
	Never		Rarely		Sometimes		Often		Always		Never		Rarely		Sometimes		Often		Always	
Checklist item	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
1. …hurts to pee.	246	69.3	62	17.5	27	7.6	14	3.9	6	1.7	908	80.6	177	15.7	35	3.1	4	0.4	2	0.2
2. …interferes with my ability to leave house.	295	82.9	29	8.1	19	5.3	9	2.5	4	1.1	1022	90.7	61	5.4	34	3.0	8	0.7	2	0.2
3. …leak when sleep.	288	80.7	28	7.8	19	5.3	14	3.9	8	2.2	921	81.8	123	10.9	52	4.6	20	1.8	10	0.9
4. … hold my pee when I need to go.	276	77.7	33	9.3	25	7.0	16	4.5	5	1.4	806	71.5	184	16.3	86	7.6	40	3.5	11	1.0
5. …leak when physically active.	291	82.0	32	9.0	21	5.9	3	0.8	8	2.3	955	84.8	101	9.0	49	4.4	16	1.4	5	0.4
6. … takes a long time to pee.	250	70.0	44	12.3	32	9.0	15	4.2	16	4.5	868	77.2	140	12.5	74	6.6	32	2.8	10	0.9
7. …leak when need to pee urgently.	263	73.9	41	11.5	30	8.4	15	4.2	7	2.0	731	65.2	177	15.8	134	12.0	54	4.8	25	2.2
8. …pee comes out like a spray.	210	59.3	52	14.7	57	16.1	20	5.6	15	4.2	412	36.6	235	20.9	263	23.4	156	13.9	60	5.3
9. …leak when cough or sneeze.	289	81.4	34	9.6	21	5.9	7	2.0	4	1.1	898	80.1	137	12.2	68	6.1	16	1.4	2	0.2
10. …hard time starting to pee.	255	72.0	53	15.0	32	9.0	10	2.8	4	1.1	889	79.1	146	13.0	63	5.6	20	1.8	6	0.5
11. …trouble emptying my bladder.	244	68.5	47	13.2	38	10.7	18	5.1	9	2.5	772	68.7	184	16.4	106	9.4	45	4.0	17	1.5
12. …urge interferes with sleep.	243	68.3	46	12.9	43	12.1	12	3.4	12	3.4	713	63.3	188	16.7	139	12.3	58	5.1	29	2.6
13. …need to push to get pee out.	225	63.4	56	15.8	37	10.4	20	5.6	17	4.8	847	75.2	175	15.5	78	6.9	17	1.5	9	0.8
14. …more pee that dribbles out.	102	28.7	49	13.8	68	19.2	60	16.9	76	21.4	544	48.4	264	23.5	192	17.1	92	8.2	33	2.9
15. …pee more often than I think I should.	226	63.7	56	15.8	39	11.0	19	5.4	15	4.2	690	61.4	205	18.2	134	11.9	66	5.9	29	2.6

*The GENDER-Q scales are copyright of McMaster University and Brigham and Women's Hospital (© 2024, McMaster University and Brigham and Women's Hospital). The GENDER-Q must not be copied, distributed, or used in any way without the prior consent of McMaster University.

eTable 6. Distribution OF Surgery, Adverse Effect Checklist*

	Face										Chest/Breast area										Genitals									
	Not at all concerned		A little concerned		Somewhat concerned		Very concerned		Extremely concerned		Not at all concerned		A little concerned		Somewhat concerned		Very concerned		Extremely concerned		Not at all concerned		A little concerned		Somewhat concerned		Very concerned		Extremely concerned	
Checklist item	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
1. Trouble urinating																					1154	77.1	181	12.1	90	6.0%	37	2.5	35	2.3
2. Trouble concentrating	230	87.5	17	6.5	10	3.8	5	1.9	1	0.4	1476	86.5	109	6.4	77	4.5	28	1.6	17	1.0	1257	84.4	121	8.1	61	4.1	29	1.9	21	1.4
3. Stiffness	212	80.9	26	9.9	14	5.3	10	3.8	0	0.0	1323	77.6	241	14.1	99	5.8	36	2.1	6	0.4	1252	84.4	160	10.8	49	3.3	17	1.1	6	0.4
4. Tenderness	195	74.7	42	16.1	14	5.4	10	3.8	0	0.0	1156	67.8	377	22.1	122	7.2	38	2.2	11	0.6	1122	75.5	227	15.3	95	6.4	27	1.8	15	1.0
5. Burning sensation	243	92.7	9	3.4	7	2.7	3	1.1	0	0.0	1556	91.3	87	5.1	39	2.3	17	1.0	5	0.3	1308	87.8	102	6.9	44	3.0	24	1.6	11	0.7
6. Odor or smell	243	93.1	5	1.9	9	3.4	2	0.8	2	0.8	1643	96.3	34	2.0	19	1.1	9	0.5	1	0.1	1002	67.0	319	21.3	115	7.7	38	2.5	21	1.4
7. Bleeding	238	90.5	15	5.7	8	3.0	2	0.8	0	0.0	1627	95.4	46	2.7	21	1.2	5	0.3	7	0.4	1289	86.4	114	7.6	56	3.8	21	1.4	12	0.8
8. Constipation	245	93.5	8	3.1	7	2.7	2	0.8	0	0.0	1592	93.5	61	3.6	29	1.7	15	0.9	5	0.3	1317	88.4	96	6.4	46	3.1	19	1.3	12	0.8
9. Throbbing feeling	234	89.7	13	5.0	10	3.8	3	1.1	1	0.4	1523	89.6	119	7.0	44	2.6	10	0.6	4	0.2	1345	90.2	104	7.0	21	1.4	13	0.9	8	0.5
10. Bruising	238	90.8	14	5.3	6	2.3	4	1.5	0	0.0	1596	93.8	60	3.5	30	1.8	15	0.9	1	0.1	1400	94.0	58	3.9	19	1.3	9	0.6	4	0.3
11. Pain when you rest	237	90.1	9	3.4	7	2.7	9	3.4	1	0.4	1480	87.0	136	8.0	53	3.1	26	1.5	7	0.4	1320	88.6	116	7.8	29	1.9	16	1.1	9	0.6
12. Discharge (eg, blood, fluid)	245	93.2	8	3.0	7	2.7	2	0.8	1	0.4	1602	94.0	58	3.4	27	1.6	13	0.8	4	0.2	1202	80.7	176	11.8	70	4.7	29	1.9	13	0.9
13. Excess perspiration (sweating)	246	93.5	6	2.3	7	2.7	3	1.1	1	0.4	1540	90.5	89	5.2	47	2.8	18	1.1	7	0.4	1324	88.8	93	6.2	52	3.5	13	0.9	9	0.6
14. Trouble sleeping	226	86.3	15	5.7	15	5.7	5	1.9	1	0.4	1432	84.1	154	9.0	70	4.1	31	1.8	15	0.9	1245	83.6	129	8.7	56	3.8	39	2.6	21	1.4
15. Tingling (pins and needles feeling)	203	77.5	36	13.7	18	6.9	5	1.9	0	0.0	1304	76.7	268	15.8	101	5.9	18	1.1	9	0.5	1280	85.9	132	8.9	54	3.6	11	0.7	13	0.9
16. Lack of feeling (numb)	163	62.0	58	22.1	28	10.6	13	4.9	1	0.4	901	52.8	518	30.4	200	11.7	57	3.3	29	1.7	1078	72.3	232	15.5	101	6.8	49	3.3	32	2.1
17. Feeling depressed or anxious	199	76.0	36	13.7	15	5.7	7	2.7	5	1.9	1349	79.2	202	11.9	95	5.6	36	2.1	22	1.3	1020	68.5	227	15.2	127	8.5	67	4.5	49	3.3
18. Feeling of pressure	224	85.5	22	8.4	10	3.8	3	1.1	3	1.1	1502	88.2	116	6.8	60	3.5	20	1.2	4	0.2	1279	85.7	126	8.4	54	3.6	18	1.2	15	1.0

19. Swelling or puffiness	211	80.5	25	9.5	17	6.5	7	2.7	2	0.8	1488	87.3	133	7.8	52	3.1	23	1.3	8	0.5	1261	84.4	135	9.0	60	4.0	25	1.7	13	0.9
20. Soreness	216	82.4	24	9.2	15	5.7	4	1.5	3	1.1	1342	79.0	241	14.2	76	4.5	28	1.6	12	0.7	1148	76.8	210	14.1	85	5.7	34	2.3	17	1.1
21. Pain when you move around	233	88.9	15	5.7	7	2.7	6	2.3	1	0.4	1451	85.4	152	8.9	55	3.2	31	1.8	11	0.6	1245	83.5	166	11.1	43	2.9	21	1.4	16	1.1
22. Pulling sensation	230	87.5	18	6.8	9	3.4	6	2.3	0	0.0	1259	73.9	305	17.9	94	5.5	36	2.1	9	0.5	1253	84.0	156	10.5	56	3.8	13	0.9	14	0.9
23. Aching feeling	233	88.6	14	5.3	8	3.0	6	2.3	2	0.8	1431	84.2	177	10.4	61	3.6	25	1.5	6	0.4	1262	84.7	135	9.1	58	3.9	21	1.4	14	0.9
24. Feeling tired	219	83.3	27	10.3	9	3.4	5	1.9	3	1.1	1398	82.2	150	8.8	98	5.8	33	1.9	22	1.3	1115	74.7	188	12.6	114	7.6	53	3.6	22	1.5
25. Itchiness	213	81.0	35	13.3	7	2.7	6	2.3	2	0.8	1346	79.1	230	13.5	91	5.3	25	1.5	9	0.5	1234	82.9	182	12.2	49	3.3	14	0.9	10	0.7
26. Tightness	212	81.5	29	11.2	14	5.4	4	1.5	1	0.4	1283	75.2	293	17.2	89	5.2	33	1.9	9	0.5	1150	77.1	198	13.3	92	6.2	31	2.1	21	1.4
27. Discomfort	215	82.4	25	9.6	13	5.0	4	1.5	4	1.5	1335	78.5	247	14.5	76	4.5	35	2.1	8	0.5	1105	74.0	242	16.2	86	5.8	39	2.6	21	1.4
28. Feeling light-headed	241	91.6	13	4.9	3	1.1	6	2.3	0	0.0	1582	93.1	62	3.6	37	2.2	12	0.7	6	0.4	1338	89.7	91	6.1	37	2.5	16	1.1	9	0.6
29. Trouble eating or drinking	238	90.5	8	3.0	12	4.6	3	1.1	2	0.8	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
30. Trouble breathing through your nose	215	82.4	28	10.7	12	4.6	4	1.5	2	0.8	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

eTable 7. Mean Scores and Key Demographic Characteristics for Scales

Scale	Rasch N	Scale score		Age				Gender identity						Preferred appearance outcome of care – overall						% Missing	Sample who completed in survey
		Mean	SD	Mean	SD	Min	Max	Man		Woman		Other		Masculinization		Feminization		Other			
								n	%	n	%	n	%	n	%	n	%	n	%		
HEALTH-RELATED QUALITY OF LIFE																					
Body Image	4525	57	24	33	12	18	83	1526	34	1610	36	1389	31	2226	49	1806	40	491	11	1	Core
Gender Dysphoria	4519	62	20	33	12	18	83	1514	34	1626	36	1379	31	2210	49	1821	40	485	11	3	Core
Social Acceptance	4621	73	17	33	12	18	83	1548	34	1659	36	1414	31	2265	49	1857	40	496	11	2	Core
Psychological Distress	4454	65	21	33	12	18	83	1499	34	1593	36	1362	31	2189	49	1787	40	476	11	1	Core
Psychological Well-Being	4557	62	20	33	12	18	83	1532	34	1626	36	1399	31	2239	49	1824	40	492	11	2	Core
Treatment Outcome	3469	75	22	34	12	18	83	1263	36	1305	38	901	26	1793	52	1428	41	246	7	2	Had surgery on face, top, bottom - asked to think of most recent
SEXUAL																					
Sexual Well-Being	3898	60	17	32	11	18	76	1369	35	1316	34	1213	31	2007	52	1480	38	410	11	4	Engaged in sexual activity in the last year; and if had bottom surgery - have engaged in sex after surgery
Orgasm	1470	65	19	37	12	18	76	463	32	790	54	217	15	558	38	867	59	44	3	3	
URINATION																					
Urinary Function - CHECKLIST																					Wanted/had bottom surgery, report trouble urinating and do not currently have a catheter
Urinary Catheter	215	50	21	36	13	19	76	52	24	145	67	18	8	58	27	156	73	1	1	3	Had bottom surgery in the last 6 months and had a catheter
GENDER PRACTICES																					
Binding - Well-Being	367	65	19	26	8	18	64	207	56	0	0	160	44	314	86	0	0	53	14	1	Reported binding in the past week
Binding - Chest Symptoms	367	69	18	26	8	18	64	208	57	0	0	159	43	314	86	0	0	53	14	1	
Binding – Skin Symptoms	366	72	19	26	8	18	64	207	57	0	0	159	43	313	86	0	0	53	15	1	
Tucking - Symptoms	306	77	18	36	13	18	73	0	0	264	86	42	14	0	0	295	96	11	4	6	Reported tucking in the past week
VOICE																					
Sound	5415	52	18	33	12	18	83	1822	34	1991	37	1602	30	2648	49	2221	41	545	10	5	Core
Distress	5367	64	21	33	12	18	83	1806	34	1975	37	1586	30	2627	49	2203	41	536	10	2	Core
HAIR																					
Face - Feminization	1584	36	16	38	14	18	81	1	0	1332	84	251	16	0	0	1584	100	0	0	5	All, except those who did not grow facial hair
Face - Masculinization	2043	58	19	30	10	18	71	1443	71	1	0	599	29	2043	100	0	0	0	0	7	
Head	1703	52	18	36	13	18	81	455	27	915	54	333	20	597	35	1008	59	97	6	3	All, except those who were bald
FACE & NECK																					
Face Overall	4898	56	19	33	12	18	83	1637	33	1790	37	1471	30	2392	49	1995	41	508	10	4	Core

Facial Features	4854	59	17	33	12	18	83	1622	33	1776	37	1456	30	2364	49	1981	41	506	10	2	Core
Upper Face	1582	52	20	36	13	18	81	149	9	1173	74	260	16	201	13	1295	82	84	5	3	Core
Eyebrows	1564	59	21	36	13	18	81	146	9	1161	74	257	16	197	13	1281	82	84	5	2	Core
Cheeks	866	51	23	38	14	18	81	85	10	634	73	147	17	116	13	696	81	52	6	3	Treatment/Surgery status - unsure, want, had, had & need revisions
Nose	1438	47	22	37	13	18	81	141	10	1012	70	285	20	202	14	1116	78	119	8	3	
Nostrils	1386	56	24	37	13	18	81	137	10	974	70	275	20	197	14	1074	78	115	8	2	
Lips	1050	49	20	39	14	18	81	92	9	766	73	192	18	134	13	835	80	80	8	4	
Chin	1496	45	24	36	13	18	81	213	14	1006	67	277	19	290	19	1113	75	91	6	2	
Jawline	1905	42	22	34	13	18	81	387	20	1031	54	487	26	568	30	1155	61	180	10	3	
Adam’s Apple	1146	69	26	37	13	18	81	1	0	1012	88	133	12	0	0	1120	98	26	2	3	
BODY																					
Body	4973	50	19	33	12	18	83	1669	34	1798	36	1506	30	2441	49	2010	40	520	11	3	Core
Buttocks	1406	42	21	36	13	18	81	294	21	756	54	356	25	436	31	839	60	131	9	3	Treatment/Surgery status - unsure, want, had, had & need revisions
Waist	2225	37	21	34	13	18	77	643	29	945	43	637	29	965	43	1048	47	212	10	2	
BREASTS																					
Breasts	2131	57	16	38	14	18	83	1	0	1855	87	275	13	0	0	2129	100	2	0	4	Surgery status - unsure, want, had, had & need revisions OR prefer fem chest and have bra cup size at least AA
Nipples & Areolas	2071	64	20	38	14	18	83	1	0	1804	87	266	13	0	0	2069	100	2	0	3	
GENITAL FEMINIZATION																					
Vagina	1236	56	21	40	14	18	83	0	0	1133	92	103	8	0	0	1221	99	15	1	6	Had feminizing bottom surgery more than 2 weeks ago
Labia	1152	60	19	40	14	18	83	0	0	1061	92	91	8	0	0	1138	99	14	1	6	Had feminizing bottom surgery more than 2 weeks ago & have inner and/or outer labia
Clitoris	1118	63	23	40	14	18	83	0	0	1029	92	89	8	0	0	1107	99	11	1	2	Had feminizing bottom surgery more than 2 weeks ago & have clitoris
Dilation	930	60	20	39	13	18	77	0	0	860	93	70	8	0	0	923	99	7	1	1	Had feminizing bottom surgery more than 2 weeks ago & use a dilator
CHEST																					
Chest	2857	66	28	29	10	18	74	1741	61	4	0	1112	39	2762	97	0	0	95	3	3	Surgery status - want, had, had & need revisions
Scars	1927	79	18	31	10	18	71	1289	67	4	0	634	33	1878	98	0	0	49	3	3	Had chest surgery more than 6 months ago
Nipples & Areolas	2535	60	22	29	10	18	74	1619	64	3	0	913	36	2479	98	0	0	56	2	2	Surgery status - want, had, had & need revisions; and if had surgery had a least one nipple
GENITAL MASCULINIZATION																					
Penis	391	62	18	36	12	19	69	340	87	0	0	51	13	381	97	0	0	10	3	4	Had masculinizing bottom surgery more than 2 weeks ago
Penis Sensation	281	55	27	36	12	19	69	333	87	0	0	49	13	374	98	0	0	8	2	8	
Glans	223	63	20	35	11	19	68	202	91	0	0	21	9	219	98	0	0	4	2	4	Had surgery to create a glans more than 2 weeks ago
Scrotum	310	60	19	36	12	19	68	282	91	0	0	28	9	307	99	0	0	3	1	5	Had surgery to create a scrotum more than 2 weeks ago

Perineum	174	63	27	35	12	19	68	158	91	0	0	16	9	173	99	0	0	1	1	5	Had surgery to create a perineum more than 2 weeks ago
Donor Site – Forearm or Thigh	252	65	22	35	11	19	69	227	90	0	0	25	10	248	98	0	0	4	2	0	Donor site located on forearm, thigh, lower leg or back, and phalloplasty not in last 2 weeks
Donor Site – Adverse Effects	251	78	18	35	11	19	69	226	90	0	0	25	10	247	98	0	0	4	2	5	Donor site located on forearm, thigh, lower leg or back, and phalloplasty not in last 2 weeks
Testicular Implants	95	57	24	39	14	19	67	90	95	0	0	5	5	95	100	0	0	0	0	4	Have testicular implants
Erectile Device	78	77	52	37	11	21	66	73	94	0	0	5	6	76	97	0	0	2	3	10	Have an erectile device
EXPERIENCE OF CARE																					
Health Professional	3017	87	18	33	12	18	81	1026	34	1167	39	824	27	1481	49	1312	44	223	7	4	Visited a healthcare professional in the last 6 months for gender-affirming care
Clinic	2333	87	18	34	12	18	81	800	34	910	39	623	27	1153	49	1018	44	162	7	3	Visited a healthcare professional in the last 6 months and physically attended a clinic or office with employees (office staff)
Surgery - Information	530	76	21	32	12	18	76	169	32	190	36	171	32	275	52	205	39	50	9	3	Had top or bottom surgery in the last 6 months
Surgery - Adverse Effects - CHECKLIST																					Had facial, top or bottom surgery
Surgery – Return to Activity	594	85	20	32	12	18	76	190	32	212	36	192	32	310	52	227	38	57	10	3	Had top or bottom surgery in the last 6 months

Data Sharing Statement

Kaur. Development and Assessment of a Patient-Reported Outcome Instrument for Gender-Affirming Care. *JAMA Netw Open*. Published April 18, 2025.
doi:10.1001/jamanetworkopen.2025.4708

Data
Data available: No