


RESEARCH

Open Access



International study to develop a patient-reported outcome measure to evaluate outcomes of gender-affirming care - the GENDER-Q

Manraj N. Kaur^{1*} , Shane D. Morrison², Shelby L. Kennedy³, Tim C. van de Grift⁴, Astrid Højgaard⁵, Amalie Lind Jacobsen⁶, Natasha Johnson³, Margriet G. Mullender⁴, Lotte Poulsen⁶, Thomas Satterwhite⁷, Richard Santucci⁸, John Semple⁹, Charlene Rae³, Kinusan Savard¹⁰, Jens Ahm Sørensen⁶, Danny Young-Afat⁴, Andrea L. Pusic¹ and Anne F. Klassen³

Abstract

Background To meaningfully understand outcomes of gender-affirming care, patient-reported outcome measures (PROMs) that are grounded in what matters to individuals seeking care are urgently needed. The objective of this study was to develop a comprehensive PROM to assess outcomes of gender-affirming care in clinical practice, research, and quality initiatives (the GENDER-Q).

Methods Internationally established guidelines for PROM development were used to create a field test version of the GENDER-Q. In-depth interviews were conducted from December 2018 to March 2020 with youth and adults aged 16 years and older who were seeking or had received gender-affirming care at outpatient clinics providing gender-affirming care located within tertiary care centers or communities in Canada, Denmark, the Netherlands or the US. Data were analyzed and used to develop a conceptual framework and an item pool, which was used to develop preliminary scales. Between February 2021 to November 2021, iterative feedback was sought from clinicians and patient participants on the scales and used to refine or develop new scales. The revised scales were pilot-tested using a crowd-sourcing platform between February 2022 and April 2022.

Results Data from interviews with 84 participants (aged 34 ± 14 years) resulted in a conceptual framework of the GENDER-Q with 13 domains measuring 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. Preliminary versions of 44 scales were developed covering most concepts in the conceptual framework. Iterative feedback was obtained from clinician experts (4 to 37 experts per scale; response rate, 67%) and 7–14 patient participants (depending on scale). All scales were refined, and 15 new scales were developed, resulting in 55 scales in the field test version of the GENDER-Q. In total, 601 transgender and gender diverse (TGD) people (aged 25 ± 6 years) participated in the pilot field test and the data were used to make changes to the field test survey.

*Correspondence:

Manraj N. Kaur
mkaur17@bwh.harvard.edu

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Conclusion The GENDER-Q was developed using extensive input from TGD individuals and clinician experts and represents the most comprehensive set of independently functioning scales that are available to date. An international field test of the GENDER-Q was completed in 2024 and the GENDER-Q is available for use in patient care, clinical research and quality improvement efforts.

Keywords Gender-affirming care, Transgender, Gender diverse, Patient-reported outcome measure, Patient-reported outcomes, Top surgery, Bottom surgery, Facial feminization, Nonbinary, Gender-affirming surgery

Background

Transgender and gender diverse (TGD) people are those whose gender identities or expressions differ from their sex assigned at birth [1, 2]. Gender affirmation refers to the process of recognizing or affirming TGD people in their gender identity or expression—socially, medically, legally, behaviourally, or a combination of these [1, 2]. Gender-affirming care is medically necessary, complex, and individualized. The person-centred nature of gender-affirming care underscores the importance of understanding the individual beyond their clinical presentation, including, their symptoms, functional status, psychosocial distress, sexual well-being, treatment goals, and experiences of healthcare. This approach is well-suited for the measurement of patient-reported outcomes.

Patient-reported outcomes (PROs) are outcomes that are reported directly by a patient without any interpretation of their response by a clinician or anyone else [3]. Fundamentally, patient-reported outcome measures (PROMs) ensure that the patient's voice is meaningfully captured in treatment-decision making and comparative effectiveness research. However, no rigorously developed and validated gender-affirming care-specific PROM currently exists. Results from five recent systematic reviews on PROMs in TGD research convergently conclude that the PROMs used in TGD research are either generic, designed for cisgender populations, or were not developed following internationally recommended guidelines for the development of PROMs [4–8]. When PROMs that do not have content validity (i.e., relevant, comprehensive, and comprehensible content) in the context of gender-affirming care are used, they fail to measure what matters to patients, and hence, cannot be used to capture the voices of TGD individuals in clinical care and research.

A PROM that is rigorously developed, validated and covers the full range of PROs relevant to gender-affirming treatments and services is urgently needed. The aim of our international study was to develop and refine the content of a comprehensive, modular PROM called the GENDER-Q for use in clinical care, clinical research, quality improvement initiatives and regulatory efforts to evaluate the PROs associated with gender-affirming interventions.

Methods and analysis

A protocol paper describing the development of the GENDER-Q has been previously published [9]. The development of GENDER-Q follows internationally established guidelines for PROM development [3, 10–12] and consists of two main steps: (1) development of a field test version of the GENDER-Q, and (2) psychometric evaluation of the GENDER-Q. This paper describes step 1 of the GENDER-Q development. Figure 1 provides an overview of the development of the GENDER-Q. Research ethics board approval was obtained from the Hamilton Integrated Research Ethics Board (Canada; coordinating site), the Medical Ethical Committee at Amsterdam University Medical Center, VUmc (The Netherlands) and Advarra (United States (US)). In Denmark, the study was included on the list of health research (exempt) within the Region of Southern Denmark.

Sample and recruitment

Individuals who identified as TGD, were 16 years or older, seeking or had received gender-affirming care, able to provide informed consent and fluent in English (Canada, the Netherlands, the US), Danish (Denmark) or Dutch (the Netherlands) were recruited from six specialized high-volume centers that provide gender-affirming care in Canada, Denmark, the Netherlands, and the US. Purposive sampling was used to recruit a sample varied by age, gender, ethnicity, type of gender-affirming treatment, and time since treatment. Eligible participants were informed of the study at their in-person routine clinic appointment, by telephone, or by email and the contact information for interested patients was shared with the site research coordinator. In the Netherlands, participants were also recruited through the TGD community groups. The study information sheet was reviewed with the potential participants and written and verbal consent was obtained. Interviews took place by phone (Canada, US) or in person—in a private clinic space (Canada, Denmark, the Netherlands) or at the participant's home (the Netherlands). Participants from Canada, the Netherlands, and the US were provided monetary compensation for their participation.

See supplementary materials for interview guides for part 1, 2 and 3. All the interviews were audio-recorded and transcribed verbatim with identifying information removed.

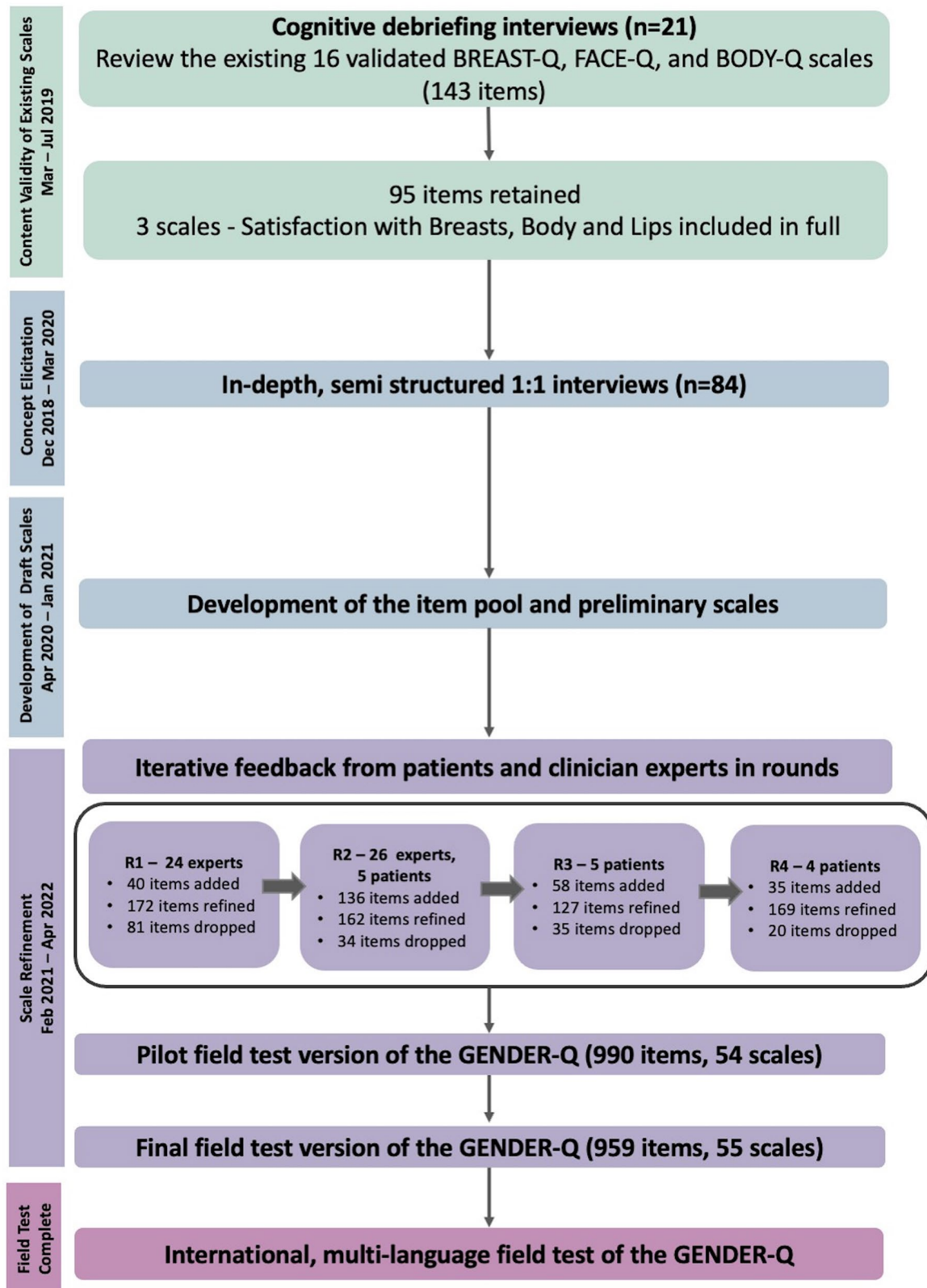


Fig. 1 Overview of the multi-step approach used for the development of the GENDER-Q

Part 1: evaluating content validity of existing scales

Cognitive debriefing interviews using the “think aloud” technique [13] were conducted with English-speaking patient participants from Canada and the US to obtain feedback on a subset of 12 scales that measure satisfaction with the body, breast, and face, and 4 scales that measure patients’ experience of care from the BREAST-Q [14], FACE-Q [15], and BODY-Q [16]. These scales were hypothesized to contain relevant content (i.e., items) for gender-affirming care (see Supplementary Material 1 for interview guide). For analysis, codes related to items from the BREAST-Q, FACE-Q and BODY-Q were organized by scale and item number and examined to identify content deemed relevant to gender-affirming care. Relevant items were included in the GENDER-Q scales.

Part 2: concept elicitation through in-depth interviews

In-depth semi-structured interviews were conducted whereby the participants were asked to describe their treatment history and any planned or future treatments. Following this, participants were asked to share the impact of gender-affirming care on their health-related quality of life (HRQL) (e.g., appearance, body image, voice, psychosocial well-being) and satisfaction with experience of care and pre-operative information (for surgical patients only) (see Supplementary Material 1 for interview guide). Demographic and clinical information was collected on a pre-set form. The part 2 interviews in Danish and Dutch were translated into English and coded alongside the English transcripts using a line-by-line approach. Codes were transferred into Excel for further analysis. Constant comparison was used to identify the top-level domains and subdomains, which led to the development of a conceptual framework covering outcomes important to people who were seeking or have had gender-affirming care. The analysis also led to the development of an item pool for use in GENDER-Q scale development. Specifically, for each important concept of interest, items from parts 1 and 2 were used to map out the concept to form a scale, and instructions, recall period, and response options were drafted. The ordering of items followed a modern psychometric approach - the Rasch Measurement Theory (RMT) - to scale design, ensuring that each scale measures a clinical hierarchy [17].

To ensure rigor, the concepts elicited during the interviews were confirmed iteratively throughout the course of interviews. Additionally, the first set of 10 interviews from each country were either double-coded by two team members who then met to review codes, or coded by one member with the codes subsequently reviewed by a second member. The study team met regularly throughout the study to review the codebook and emerging concepts.

Part 3: scale refinement

Scales were refined iteratively through multiple rounds of feedback from patient participants and clinical and research experts in gender-affirming care. Clinical experts known to the GENDER-Q team were invited via email to review the GENDER-Q scales and provide feedback. Given the large number GENDER-Q scales, experts were invited to review as many as they were able to or to focus on the subset that fell within their area of expertise. Experts were asked to use track changes or the comment feature in Microsoft Word to provide feedback on the comprehensibility, comprehensiveness and relevance of scale instructions, recall period, response options and items, and to suggest missing items.

English-speaking patient participants from the US and Canada who took part in part 2—concept elicitation interviews and new patients were invited to review the GENDER-Q scales in cognitive debriefing interviews using the think-aloud technique [13]. The interviews were conducted in three rounds by phone or an institutionally licensed virtual conferencing platform (i.e., Zoom). The patient participants were asked to comment on scale’s instructions, recall period, response options and items to ensure that the scales were easy to understand and relevant (see Supplementary Material 1 for interview guide). At the end of the interview, participants were asked if the scales and the GENDER-Q conceptual framework were comprehensive and to nominate missing items or scales [10–12]. The interviews were coded line-by-line and analyzed. Expert and patient participant input was used to iteratively refine the GENDER-Q scales until no more changes were deemed necessary.

Part 4: pilot field test

A pilot field test of the GENDER-Q scales was conducted in 2 parts using an online crowdsourcing research platform called Prolific (<https://www.prolific.co>) (Fig. 2). Prolific members who were 18 years or older, identified as transgender or nonbinary and were fluent in English (any country) were sent a link to an online REDCap survey. Participants self-selected as eligible and provided consent before completing clinical and demographic questions and the GENDER-Q scales. Branching logic was used to ensure participants only answered scales relevant to their experience. An open-ended text box was included for additional feedback. Participants were compensated at a prorated hourly rate of \$18 USD.

Exploratory Rasch Measurement Theory (RMT) analysis was performed to examine scale performance using RUMM2030 software with the unrestricted Rasch model for polytomous scales (RUMM version 2030, RUMM Laboratory Pty Ltd, Duncraig, Western Australia, Australia, 1998–2023). The analysis examined the fit of items to the Rasch model. Items with extreme misfit to the Rasch

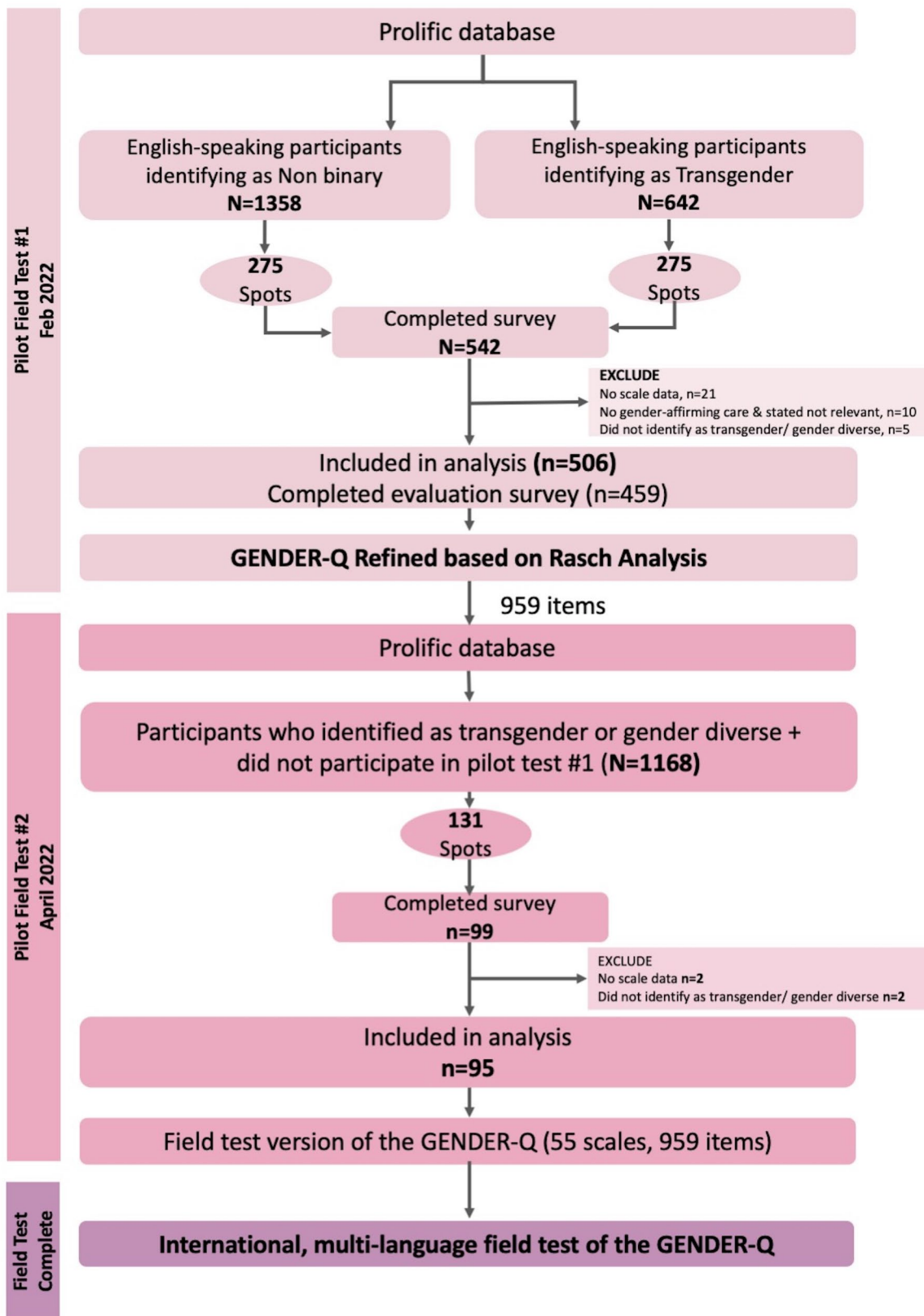


Fig. 2 Overview of the pilot field test of the GENDER-Q

model were removed. Items were reordered according to the item location order for each scale (i.e., clinical hierarchy). Open-text comments were reviewed for feedback on questions or branching logic, and the survey was updated accordingly. A second and final pilot field test was conducted with the updated version of the GENDER-Q to ensure the functionality of the survey, resulting in the final field test version of the GENDER-Q.

Results

Part 1: evaluating content validity of existing scales

Cognitive debriefing interviews ($n=21$) took place between March and July 2019. Table 1 shows the characteristics of the sample. Overall, much of the content from the existing scales resonated with participants. Table 2 shows the existing PROM scales that were included in the GENDER-Q. Of the 16 scales reviewed, 95 of the 143 items were considered relevant and covered important issues for people undergoing breast/chest surgery, body contouring and facial feminization. A further eight items were added to address missing concepts suggested by participants during the scale refinement phase (e.g., “how attractive your cheeks look?” from the FACE-Q Cheeks scale). The item set for three scales measuring satisfaction with breasts, body and lips were included in full.

Part 2: concept elicitation through in-depth interviews

A total of 85 participants were interviewed between December 2018 and March 2020. One participant subsequently withdrew from the study. Table 1 shows the characteristics of the 84 participants in the study sample. More than half of the participants were seeking or had undergone masculinization treatments. The most common procedure was chest surgery, and the least common procedures were body contouring and tracheal shaving. The analysis led to the development of a preliminary framework of concepts that are important to measure when assessing outcomes of gender-affirming care.

Part 3: scale refinement

Important concepts from the qualitative data, supplemented by relevant items from the BREAST-Q, FACE-Q and BODY-Q scales from part 1, were used to develop version 1 of the GENDER-Q. For surgical scars (chest, donor site), concepts from the qualitative data that overlapped with the SCAR-Q developed by our team were used [18, 19]. Additionally, the Animation Deformity scale from the BREAST-Q that measures the distortion in the chest appearance with the contraction of the pectoralis muscle for individuals with chest augmentation was included [20].

The GENDER-Q scales were refined between February and November 2021. We obtained feedback from 50 of 75 invited experts (response rate, 67%). The experts were

from the US ($n=28$), Denmark ($n=6$), Canada ($n=5$), The Netherlands ($n=4$), Belgium ($n=3$), Spain ($n=3$), and Switzerland ($n=1$). Experts' specialties included plastic surgery ($n=30$), psychology ($n=4$), urology ($n=3$), endocrinology ($n=3$), speech therapy ($n=3$), research ($n=3$), gynecology ($n=2$), and others ($n=2$). Depending on the scale, feedback was obtained from 4 to 37 experts and 7–14 patient participants (See Supplementary Table 2). All but one patient participant in the scale review had previously participated in the part 2 concept elicitation interview. Table 3 summarizes changes made to the GENDER-Q scales between the rounds. Items were added or deleted due to issues with comprehension including participants not interpreting key concepts and questions as intended, requesting clarification on the meaning of specific words or entire item, or experiencing difficulty with recall or judgement. Other reasons for refining or deleting items included challenges in forming responses, indications that an item was overtly sensitive or caused gender dysphoria, age or gender-related nuances that could influence interpretation or response to an item, and participants identifying items as being too long or difficult to understand.

Part 4: pilot field test

A total of 601 participants from 30 countries were included in the pilot field test (parts 1 and 2) (Fig. 2). Sample demographics for the pilot field tests are shown in Table 4.

Based on the exploratory RMT analysis of the part 1 pilot field test data, 17 items with poor item fit in 11 scales were removed. The Appearance—Face scale was divided into two scales that measured the appearance of the face and the appearance of facial parts. The Appearance—Upper Face scale was split to measure the appearance of the upper face and the appearance of eyebrows. The masculine and feminine Appearance—Facial Hair scales were modified to create a single satisfaction scale that was applicable across the gender spectrum. Seventeen GENDER-Q scales (all genital surgery-related, Return to Activity, Catheter, and Breast—Animation Deformity, and Information) did not have sufficient data for RMT analysis and were not altered. The items in all remaining scales were reordered according to the clinical hierarchy based on the item locations from the RMT analysis. Figure 3 shows the conceptual framework for GENDER-Q. Each component of the framework corresponds to an independently functioning GENDER-Q scale.

Several iterative changes were made to the GENDER-Q survey based on participants' open-text comments. First, at the start of the survey, a box was added with the following content warning: “The GENDER-Q asks detailed questions about how you look and feel, and about

Table 1 Characteristics of participants in Step 1 of GENDER-Q development

		Content generation		Part 3
		Part 1	Part 2	Scale re-
		Content validity of existing scales (<i>n</i> = 21)	Concept elicitation (<i>n</i> = 84)	finement (<i>n</i> = 14)
		N (%)	N (%)	N (%)
Country	Canada	5 (24)	20 (24)	1 (7)
	Denmark	0	12 (14)	0
	The Netherlands	0	21 (25)	0
	United States	16 (76)	31 (37)	13 (93)
Gender identity	Trans masculine	3 (14)	42 (50)	7 (50)
	Trans feminine	18 (86)	37 (44)	7 (50)
	Nonbinary / Gender queer/ Gender non-conforming	0	5 (6)	0
Age	16–19	0	14 (17)	0
	20–29	4 (19)	23 (27)	0
	30–39	9 (43)	20 (24)	8 (57)
	40–49	5 (24)	11 (13)	2 (14)
	50–59	2 (10)	11 (13)	3 (21)
	≥ 60	1 (5)	5 (6)	1 (7)
Race	White	16 (76)	53 (63)	12 (86)
	Other	4 (19)	9 (11)	2 (14)
	Prefer to not answer/missing	1 (5)	22 (26)	0
Marital status	Single, never married	5 (24)	35 (42)	3 (21)
	Married / Living common law	5 (24)	20 (24)	7 (5)
	Divorced/Separated/Not in relationship	3 (14)	10 (12)	1 (7)
	Currently in relationship	8 (38)	19 (23)	3 (21)
Education	Some high school/Completed high school	2 (10)	32 (38)	0
	Some college, trade or university	4 (19)	8 (10)	3 (21)
	Completed college, trade or university	12 (57)	36 (43)	8 (57)
	Completed Masters/Doctoral degree	3 (14)	8 (10)	3 (21)
Treatment type	Masculinization	3 (14)	47 (56)	7 (50)
	Feminization	18 (86)	37 (44)	7 (50)
Reported having	Voice surgery and/or therapy (all)	6 (29)	15 (18)	3 (21)
	Body contouring	1 (6)	1 (1)	1 (7)
	<i>Feminizing procedures</i>			
	Tracheal shave (feminization participants only)	1 (6)	3 (8)	0
	Facial feminization surgery	3 (17)	6 (16)	1 (14)
	Surgery to augment the chest	8 (44)	10 (27)	2 (29)
	Vaginoplasty	16 (89)	22 (59)	6 (86)
	<i>Masculinizing procedures</i>			
	Surgery to flatten or contour the chest	3 (100)	31 (83)	7 (50)
	Phalloplasty	0	10 (21)	5 (71)
	Metoidioplasty	0	6 (13)	1 (14)
	Scrotoplasty	0	5 (11)	2 (29)
	Glansplasty	0	5 (11)	2 (29)
	Erectile device	0	3 (6)	1 (14)

gender-affirming care. Some people have said that some parts of the GENDER-Q made them feel uncomfortable or upset. If you take part in our study, most questions are set up to allow you to skip them if you do not want to answer. Some questions are required as these are used to make sure you are asked questions that are relevant to you. In addition, based on feedback about emotionally triggering parts of the survey, the order in which the GENDER-Q scales appear in the survey was changed

with the HRQL scales, including Gender Dysphoria, moved closer to the end of the survey. Participant feedback informed changes to the clinical and demographic questions, including the addition of response options (i.e., I am not sure, prefer not to answer), rewording of hard-to-understand or answer questions (e.g., “Have you ever taken hormones or medication for the purpose of gender affirmation” was changed to “Have you taken hormones or medication in the past that you are no longer

Table 2 Scales from the BREAST-Q, FACE-Q and BODY-Q reviewed by participants and summary of change in number of items after participant feedback

PROM	Scales	No. of items in the scale	No. of items after Part 1	No. of items in the pilot field test version	Example item
BREAST-Q	Breasts	15	15	15	How your bras fit?
FACE-Q	Face overall	10	5	5	How your face looks in photos?
	Forehead & Eyebrows	6	1	3	The position of your eyebrows?
	Eyes	7	2	2	How open your eyes look?
	Lips	10	10	10	How full your lower lip looks?
	Chin	10	8	8	The size of your chin?
	Cheeks	5	3	4	How attractive your cheeks look?
	Cheekbones	10	0	2	How high your cheekbones look?
	Nose	10	9	9	The overall size of your nose?
	Nostrils	5	2	4	The shape of your nostrils?
	Lower Face & Jawline	5	2	3	The shape of your jawline?
BODY-Q	Body	10	10	10	How your clothes fit your body?
	Information	10	4	4	How the surgery would be done?
	Surgeon	10	9	9	Spent enough time with you?
	Medical team	10	8	8	Treated you with respect?
	Office staff	10	7	7	Answered all your questions?
Total		143	95	103	

taking for the purpose of gender affirmation”), providing definitions (e.g., for inner and outer labia), improving inclusivity (e.g., “chest masculinization” was changed to “chest surgery”), and addition of questions to improve clarity and interpretation of results (e.g., a question about the importance of facial hair and the concern about Adam’s apple).

Discussion

The field test version of the GENDER-Q comprises 55 independently functioning scales, covering a broad range of concepts across 13 domains relevant to individuals seeking gender-affirming care—HRQL, sexual, urination, gender practices, voice, hair, face and neck, body, breasts, genital feminization, chest, genital masculinization and experience of care. The scales were meticulously designed to enhance patient-centered shared decision-making, advance comparative effectiveness research, and support value-based gender-affirming care.

The GENDER-Q responds effectively to the demand for a rigorous, validated gender-affirming care-specific PROM by the clinical and academic communities [1, 4–8, 18]. The GENDER-Q addresses the limitations of existing PROMs used in TGD research through its adherence to internationally established guidelines for PROM development. The life stories of 84 TGD participants from four countries with different politico-legal and health-care environments were used to create the GENDER-Q scales to ensure that the scales measured important and relevant outcomes. Additionally, the development

process included a large, internationally recognized group of clinical experts in gender-affirming care, several of whom self-identified as TGD. The GENDER-Q program of research embodies a collaborative effort among lived experience experts, clinicians, HRQL researchers and PROM developers, drawing on both experiential and empirical knowledge to enhance its relevance and applicability.

Several pragmatic considerations in the development of the GENDER-Q warrant discussion. The cognitive debriefing interviews conducted to refine the scales were limited to English-speaking participants in Canada and the US. This decision was made to conserve resources related to translating the draft scales into Dutch and Danish and to avoid the need for translating the cognitive debriefing interviews to English for analysis, particularly given that the interviews were conducted in rounds. Further, depending on the scale, 7–14 participants reviewed the draft scales. While this number of patient participants may appear imbalanced compared to the number of clinical experts (ranging from 4 to 37 per scale) providing feedback, the sample size met the recommended sample size criteria for content validity established by the COSMIN guideline [10]. The revisions were made to the GENDER-Q scales in 4 iterative rounds, and it was not feasible to track the reasons for item deletion, addition or revision due to the large number of items tested. Additionally, the sample in part 1 and 2 included few individuals who identified as nonbinary. To enhance the relevance of the GENDER-Q scales across the gender spectrum, we

Table 3 Changes made to GENDER-Q scales in each round of patient and expert feedback in Part 3 and pilot field test in Part 4

Domain	GENDER-Q Scales					Round 1					Round 2				
	V1	Retain	Revise	Drop	Add	V2	Retain	Revise	Drop	Add	Retain	Revise	Drop	Add	
HRQL	Body Image	15	0	8	7	3	11	10	1	0	2				
	Gender Dysphoria	16	0	16	0	1	17	17	0	0	0				
	Social Acceptance	24	20	4	0	1	25	19	4	2	0				
	Psychological Distress	20	20	0	0	0	20	19	1	0	1				
	Psychological Well-Being	21	19	0	2	0	19	19	0	0	0				
	Treatment Outcome	22	9	10	3	0	19	15	2	2	1				
	Sexual Well-Being	21	8	9	4	2	19	8	6	5	1				
	Orgasm	-	-	-	-	-	-	-	-	-	-				
	Urination	19	18	1	0	1	20	13	6	1	3				
	Urinary Catheter	16	14	1	1	0	15	14	1	0	1				
Gender Practices	Binding Well-Being	-	-	-	-	-	13	12	1	0	0				
	Binding: Adverse Effects	25	18	2	5	0	20	19	1	0	0				
	Tucking: Adverse Effects	15	6	5	4	2	13	13	0	0	2				
	Sound	16	7	7	2	3	17	14	3	0	4				
Voice	Distress	14	10	4	0	0	14	11	2	1	2				
	Hair – Face [§]	14	12	0	2	2	14	12	1	1	2				
	Hair – Face	-	-	-	-	-	13	7	4	2	5				
	Hair - Head	14	10	2	2	0	12	11	1	0	3				
Face & Neck	Hair – Body	13	0	0	0	0	0	-	-	-	-				
	Face	19	18	1	0	3	22	21	1	0	3				
	Upper Face	18	12	2	4	0	14	11	3	0	1				
	Nose	17	14	2	1	0	16	7	9	0	4				
Body	Nostrils	7	6	1	0	0	7	5	2	0	2				
	Lips	20	16	2	2	1	19	17	2	0	3				
	Cheeks**	-	-	-	-	-	10	8	2	0	3				
	Chin	14	10	2	2	0	12	10	1	1	3				
	Jawline	12	9	2	1	0	11	6	4	1	4				
	Adam's Apple	12	10	1	1	0	11	10	0	1	4				
	Body**	-	-	-	-	-	11	10	1	0	1				
	Buttocks**	-	-	-	-	-	7	6	1	0	3				
	Waist**	-	-	-	-	-	8	7	1	0	1				
	Breasts	21	18	3	0	0	21	19	2	0	1				
Breasts	Nipples Areolas	11	2	8	1	2	12	9	3	0	1				
	Animation Deformity**	-	-	-	-	-	12	12	0	0	0				
	Vagina	17	13	4	0	1	18	12	5	1	4				
	Labia	30	25	3	2	1	29	12	16	1	0				
Genital Feminization	Clitoris	16	9	5	2	0	14	6	2	6	8				

Table 3 (continued)

Domain	GENDER-Q Scales	V1			V2			Round 1			Round 2			
		Retain	Revise	Drop	Add	Retain	Revise	Drop	Add	Retain	Revise	Drop	Add	
Chest		21	14	6	1	16	13	3	0	1				
	Nipples/areolas	15	8	3	0	12	6	6	0	1				
	Scars**	-	-	-	-	12	12	0	0	6				
	Penis	24	6	2	4	26	12	13	1	5				
	Penis Sensation	10	0	0	2	12	12	0	0	5				
	Glans	15	3	4	2	13	0	12	1	1				
	Scrotum	22	3	4	0	18	10	8	0	6				
	Perineum	12	3	2	1	11	2	9	0	2				
	Phalloplasty Flap	14	0	0	0	14	8	5	1	1				
	Phalloplasty Scars**	-	-	-	-	12	12	0	0	5				
	Donor Site: Adverse Effects	10	0	0	1	11	9	2	0	2				
	Testicular implants	11	2	0	1	12	4	8	0	3				
	Erectile device	17	13	1	0	16	6	5	5	8				
	Healthcare Professional	40	3	7	2	35	35	0	0	2				
	Clinic	21	6	3	1	19	18	0	1	1				
	Surgery - Information	20	6	1	0	19	19	0	0	5				
	Surgery - Adverse Effects	25	0	0	1	26	26	0	0	6				
	Surgery - Return to Activity	16	0	0	0	17	15	2	0	3				
	Domain	GENDER-Q Scales	Round 3			V4			Round 4			Pilot FT		
			Retain	Revise	Drop	Add	Retain	Revise	Drop	Add	Retain	Revise	Drop	Add
HRQL	Body Image	12	1	0	1	14	13	0	0	14				
	Gender	14	3	0	1	18	0	0	0	18			2	
Dysphoria	Social	15	6	2	1	22	21	1	0	21			1	
	Acceptance													
Psychological Distress	Psychological	21	0	0	1	22	20	2	0	20			1	
	Well-Being	19	0	0	2	21	21	0	0	21				
Treatment Outcome	Treatment	16	1	1	0	17	16	1	0	16				
	Sexual Well-Being	15	2	1	3	17	16	0	1	18			2	
Sexual	Orgasm	11	3	0	1	15	0	4	4	15			1	
	Urination	16	4	2	5	25	18	1	1	25				
Urinary	Urinary Catheter	9	7	0	0	16	11	4	1	15			NA	
	Binding	13	-	-	-	13	7	1	0	12			NA	
Practices	Well-Being													
	Binding: Adverse Effects	20	-	-	-	20	20	0	1	21				

Table 3 (continued)

Domain	GENDER-Q Scales	V3			V4			Round 4			Pilot FT	
		Retain	Revise	Drop	Add	Drop	Retain	Revise	Drop	Add	Drop	Revise
Voice	Tucking:Adverse Effects	15	3	2	2	15	14	1	0	0	15	
	Sound	21	2	1	1	21	17	3	1	2	22	3
Hair	Distress	15	1	1	0	14	11	2	1	1	14	1
	Hair – Face [§]	15	3	1	3	17	11	6	0	1	18	Merged into 1 scale
	Hair – Face	16	–	–	–	16	16	0	0	1	17	
	Hair - Head	15	4	2	4	17	14	3	0	1	18	Scale dropped
Face & Neck	Hair – Body	–	–	–	–	–	–	–	–	–	–	Split into 2 scales
	Face	25	2	0	3	28	26	2	0	2	30	Split into 2 scales
Body	Upper Face	15	6	1	7	21	19	1	1	1	21	
	Nose	20	4	2	0	18	13	4	1	1	18	
	Nostrils	9	1	2	2	9	6	3	0	1	10	
	Lips	22	4	0	2	24	16	8	0	0	24	
	Cheeks**	13	3	0	2	15	15	0	0	0	15	
	Chin	14	3	0	0	14	12	2	0	0	14	1
	Jawline	14	3	1	1	14	13	1	0	0	14	2
	Adam's Apple	14	2	2	2	14	13	1	0	2	16	
	Body**	12	2	0	0	12	12	0	0	0	12	
	Buttocks**	10	1	0	0	10	9	1	0	1	11	
Breasts	Waist**	9	1	1	0	8	7	1	0	0	8	
	Breasts	22	3	0	0	22	21	1	0	1	23	
	Nipples Areolas	13	3	2	2	13	13	0	0	0	13	
Genital Feminization	Animation Deformity**	12	0	0	0	12	12	0	0	0	12	NA
	Vagina	21	3	2	0	19	16	3	0	2	21	NA
Chest	Labia	28	2	1	1	28	24	3	1	1	28	NA
	Clitoris	16	3	1	2	17	14	3	0	0	17	NA
	Chest	17	–	–	–	17	16	1	0	0	17	NA
Genital Masculinization	Nipples/areolas	13	3	1	2	14	12	2	0	1	15	
	Scars**	18	–	–	–	18	17	1	0	1	19	
	Penis	30	–	–	–	30	21	9	0	3	33	NA
Glans	Penis Sensation	17	–	–	–	17	0	15	2	1	16	NA
	Glans	13	–	–	–	13	9	4	0	2	15	NA

Table 3 (continued)

Domain	GENDER-Q Scales	V3				V4				Round 4				Pilot FT	
		Retain	Revise	Drop	Add	Retain	Revise	Drop	Add	Retain	Revise	Drop	Add	Drop	Revise
Experience of care	Scrotum	24	-	-	-	24	19	5	0	1	25	NA	NA	NA	NA
	Perineum	13	3	0	2	15	10	4	1	0	14	NA	NA	NA	NA
	Phalloplasty Flap	14	-	-	-	14	13	1	0	0	14	NA	NA	NA	NA
	Phalloplasty Scars**	17	-	-	-	17	17	0	0	1	18	NA	NA	NA	NA
	Donor Site: Adverse Effects	13	-	-	-	13	12	1	0	0	13	NA	NA	NA	NA
	Testicular implants	15	-	-	-	15	10	4	1	0	14	NA	NA	NA	NA
	Erectile device	19	-	-	-	19	15	4	0	0	19	NA	NA	NA	NA
	Healthcare Professional	37	29	6	2	35	32	3	0	0	35	2	NA	NA	NA
	Clinic	19	3	16	0	20	17	3	0	0	20	1	NA	NA	NA
	Surgery - Information	24	12	11	1	26	18	8	0	0	26	NA	NA	NA	NA
	Surgery - Adverse Effects	32	28	1	3	30	25	5	0	0	30	0	0	0	0
	Surgery - Return to Activity	20	19	1	0	20	16	4	0	0	20	NA	NA	NA	NA

FT field test, NA not applicable due to the insufficient sample size of the pilot field test sample

**Represents existing BREAST-Q, FACE-Q, BODY-Q or SCAR-Q scales

‡Feminization version of the scale

Table 4 Demographic characteristics of the pilot field test sample

		Pilot field test 1		Pilot field test 2		Total	
		N=506		N=95		N=601	
		n	%	n	%	n	%
Gender identity (best describes)	Trans masculine	137	27	56	59	193	32
	Trans feminine	81	16	15	16	96	16
	Nonbinary	258	51	19	20	277	46
	Other (e.g., gender fluid)	30	6	5	5	35	6
Sex assigned on original birth certificate	Male	145	29	22	23	167	28
	Female	361	71	73	77	434	72
Age	18–19 years	60	12	18	19	78	13
	20–29 years	365	72	59	62	424	71
	30–39 years	67	13	14	15	81	13
	40 years or more	14	3	4	4	18	3
Race	White	408	81	75	79	483	80
	Latin American (e.g., Mexico, Central America, Caribbean islands)	46	9	6	6	52	9
	Black (e.g., African, Caribbean, African Canadian/American descent)	25	5	2	2	27	4
	Other	27	5	12	13	39	6
Highest level of education attained	Some high school	18	4	3	3	21	3
	Completed high school	98	19	25	26	123	20
	Some college or trade school or university	203	40	33	35	236	39
	Completed college or trade school or university degree	132	26	23	24	155	26
	Some Masters or Doctoral degree	22	4	4	4	26	4
	Completed Masters or Doctoral degree	31	6	6	6	37	6
Country of residence	Prefer not to answer	2	0	1	1	3	0
	United States	145	29	29	31	174	29
	United Kingdom	104	21	29	31	133	22
	Poland	41	8	5	5	46	8
	Mexico	31	6	2	2	33	5
	Canada	23	5	3	3	26	4
	Italy	22	4	0	0	22	4
	Portugal	19	4	7	7	26	4
	South Africa	19	4	3	3	22	4
	Spain	19	4	2	2	21	3
	Germany	14	3	4	4	18	3
	Netherlands	10	2	0	0	10	2
	Other	59	12	11	12	70	12
	Reported having	Hormones (currently)	185	37	51	54	236
Hair transplant (head)		3	1	0	0	3	0
Scalp Advancement		1	0	1	1	2	0
Facial surgery		11	2	3	3	14	2
Facial hair removal (feminization participants only)*		26	20	6	35	32	22
Voice therapy		48	10	7	7	55	9
Voice surgery		1	0	0	0	1	0
Tracheal shave (feminization participants only)®		4	3	1	5	5	3
Surgery to flatten or contour the chest (masculinization participants only)®		60	20	17	24	77	21
Surgery to augment the chest (feminization participants only)®		7	5	1	4	8	5
Surgery to create genitals		9	2	2	2	11	2
Body contouring		5	1	1	1	6	1

*Pilot field test 1, n=129; Pilot field test 2, n=17

®Pilot field test 1, n=145; Pilot field test 2, n=22

#Pilot field test 1, n=296; Pilot field test 2, n=70

^Pilot field test 1, n=138; Pilot field test 2, n=25



Fig. 3 Conceptual Framework of the GENDER-Q

implemented sampling quotas to recruit a larger number of nonbinary participants for the pilot field test. Other limitations of the pilot field test include a lack of diversity in age, race and educational attainment and a lack of participants who had undergone genital gender-affirming surgery. This last limitation hindered our ability to

explore the psychometric performance of certain scales prior to the international field test. The international, multi-language field test (step 2) of the GENDER-Q (completed in 2024) overcomes these limitations through the recruitment of a large and diverse international sample from clinical and community settings.

Conclusions

The GENDER-Q represents the most comprehensive set of scales that are specific to measuring outcomes of gender-affirming care. The international field test was completed in 2024. The GENDER-Q is available for use in patient care, clinical research and quality improvement efforts through www.qportfolio.org.

Abbreviations

HRQL	Health-related quality of life
COSMIN	Consensus-based Standards for the selection of health Measurement Instruments
PRO	Patient-reported outcome
PROM	Patient-reported outcome measure
RMT	Rasch Measurement Theory
TGD	Transgender and gender diverse
US	United States
USD	United States Dollar

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s41687-024-00785-x>.

Supplementary Material 1

Acknowledgements

We are grateful to all our study participants who shared their life stories with us and participated in the pilot test to help us create the GENDER-Q. We are also thankful to the clinical experts who provided their input on the preliminary drafts of the scale.

Author contributions

AK, MK, AP: conception, design, acquisition, analysis and interpretation of data (all phase 1), creation of the GENDER-Q, and drafted the manuscript; SDM, KS: interpretation of data; SK: analysis and interpretation of data; TCG, AH, ALJ, NJ, MM, LP, TS, RS, JS, JAS, DYA: acquisition of participants; CR: acquisition, analysis and interpretation of pilot field test data. All authors reviewed and provided feedback on the pre-field test versions of the GENDER-Q. All authors have approved the submitted version of the manuscript.

Funding

Phase 1 of this study is funded by research grants received from the Canadian Institutes for Health Research (CIHR) Foundation Grant (Award #FDN 167288), CIHR Strategy for Patient-Oriented Research (SPOR)-Patient-Oriented Research Collaboration Grant (Award #157062), the Plastic Surgery Foundation National Endowment for Plastic Surgery Grant (Award #568801) and ZonMw (Project number #636330001). The authors have no financial interest to declare in relation to the content of this article.

Data availability

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

Research ethics board approval was obtained from the Hamilton Integrated Ethics Board (Canada; coordinating site), the Medical Ethical Committee at Amsterdam University Medical Center, VUmc (The Netherlands) and Advarra (United States (US)). In Denmark, the study was included on the list of health research (exempt) within the Region of Southern Denmark. All participants provided written and verbal (for qualitative interviews only) to participate in the study.

Consent for publication

Not applicable.

Competing interests

The GENDER-Q was co-developed by Drs Anne Klassen, Manraj Kaur, and Andrea Pusic; McMaster University and Mass General Brigham own the copyright. Drs Klassen, Kaur and Pusic will receive a share of license revenues as royalties for its use in for-profit research based on their institution's inventor sharing policy. Klassen provides research consulting services to the pharmaceutical industry through EVENTUM Research (Hamilton, Ontario, Canada).

Author details

¹Brigham and Women's Hospital, Harvard Medical School, 75 Francis S, Boston, MA 02116, USA

²Division of Plastic Surgery, Department of Surgery, University of Washington, 325 9th Ave, Mail Stop #359796, Seattle, WA 98104, USA

³Department of Pediatrics, McMaster University, 1280 Main Street W, Hamilton, ON L8N 3Z5, Canada

⁴Department of Plastic Reconstructive and Hand Surgery, Center of Expertise on Gender Dysphoria, Amsterdam Public Health Research Institute, VU University Medical Center, De Boelelaan 1117, Amsterdam 1081 HV, The Netherlands

⁵Sexological Center, Center for Gender, Center for Rape Victims, Aalborg University Hospital, Aalborg, Denmark

⁶Research Unit for Plastic Surgery, University of Southern Denmark and Odense, Denmark and OPEN, Odense Explorative Patient Network, Odense, Denmark

⁷Align Surgical Associates Inc., 2299 Post St. Suite 207, San Francisco, CA 94115, USA

⁸Crane Center, 4407 Bee Caves Rd. Ste. 612, Austin, TX 78746, USA

⁹Division of Plastic, Reconstructive and Aesthetic Surgery, University of Toronto, The Rotman/Stewart Building, 149 College Street, 5th Floor, Suite 508, Toronto, ON M5T 1P5, Canada

¹⁰Fleming College - Emeritus, RR1, Noelville, ON P0M 2N0, Canada

Received: 2 February 2024 / Accepted: 30 August 2024

Published online: 19 November 2024

References

1. Coleman E, Radix AE, Bouman WP et al (2022) Standards of care for the health of transgender and gender diverse people, version 8. *Int J Transgend Health* 23(sup1):S1–259
2. Reisner SL, Poteat T, Keatley J et al (2016) Global health burden and needs of transgender populations: a review. *Lancet* 388(10042):412–436
3. Patient-Reported Outcome Measures (2009) Use in medical product development to support labeling claims—guidance for industry. The U.S. Food and Drug Administration. <https://www.fda.gov/media/77832/download>. Accessed 23 Aug 2022
4. Oles N, Darrach H, Landford W et al (2022) 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* 275(1):e52–e66
5. Oles N, Darrach H, Landford W et al (2022) 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* 275(1):e67–e74
6. Dy GW, Nolan IT, Hotaling J, Myers JB (2019) Patient reported outcome measures and quality of life assessment in genital gender confirming surgery. *Transl Androl Urol* 8(3):228–240
7. Andréasson M, Georgas K, Elander A, Selvaggi G (2018) Patient-reported outcome measures used in gender confirmation surgery: a systematic review. *Plast Reconstr Surg* 141(4):1026–1039
8. Barone M, Cogliandro A, Di Stefano N, Tambone V, Persichetti P (2017) A systematic review of patient-reported outcome measures following transsexual surgery. *Aesthetic Plast Surg* 41(3):700–713
9. Klassen AF, Kaur M, Johnson N et al (2018) 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* 8(10):e025435
10. Mokkink LB, Terwee CB, Patrick DL et al (2010) The COSMIN checklist for assessing the methodological quality of studies on measurement properties

- of health status measurement instruments: an international Delphi study. *Qual Life Res* 19(4):539–549
11. Patrick DL, Burke LB, Gwaltney CJ et al (2011) Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: part 1—eliciting concepts for a new PRO instrument. *Value Health* 14(8):967–977
 12. Patrick DL, Burke LB, Gwaltney CJ et al (2011) Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research practices Task Force report: part 2—assessing respondent understanding. *Value Health* 14(8):978–988
 13. Willis GB (2004) *Cognitive interviewing: a tool for improving questionnaire design*. Sage
 14. Pusic AL, Klassen AF, Scott AM, Klok JA, Cordeiro PG, Cano SJ (2009) Development of a new patient-reported outcome measure for breast surgery: the BREAST-Q. *Plast Reconstr Surg* 124(2):345–353
 15. Klassen AF, Cano SJ, Scott A, Snell L, Pusic AL (2010) Measuring patient-reported outcomes in facial aesthetic patients: development of the FACE-Q. *Facial Plast Surg* 26(4):303–309
 16. Klassen AF, Cano SJ, Alderman A et al (2016) The BODY-Q: a patient-reported outcome instrument for weight loss and body contouring treatments. *Plast Reconstr Surg Glob Open* 4(4):e679
 17. Rasch G (1960) Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests
 18. Ziolkowski NI, Pusic AL, Fish JS et al (2020) Psychometric findings for the SCAR-Q patient-reported outcome measure based on 731 children and adults with surgical, traumatic, and burn scars from four countries. *Plast Reconstr Surg* 146(3):331e–8e
 19. Klassen AF, Ziolkowski N, Mundy LR et al (2018) Development of a new patient-reported outcome instrument to evaluate treatments for scars: the SCAR-Q. *Plast Reconstr Surg Glob* 6(4)
 20. Tsangaris E, Pusic AL, Kaur MN et al (2021) Development and psychometric validation of the BREAST-Q animation deformity scale for women undergoing an Implant-based breast Reconstruction after Mastectomy. *Ann Surg Oncol* 28(9):5183–5193

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary material 1. SEMI-STRUCTURED INTERVIEW GUIDES

Part 1. Evaluating content validity of existing scales

Part 3. Scale refinement

Instructions

- What are the instructions asking you to do? Please explain in your own words.
- Is the time frame for responding to the scale appropriate?
- Are there any words we should change to make the instructions easier?

Items

- In your own words, what is this item asking?
- What do you think of when answering this item?
- Are any words difficult to understand or offensive?
- Was this item hard or easy to answer and why?
- Does this item measure an important issue for you?
- What do you think about the response choices?

At the end of each scale

- In your own words, what is this group of items asking about?
- Does this group of items measure an important issue for you?
- Are there any items that do not belong with the rest?
- Thinking about that group of items, what are we missing?

At the end of the interview

- What are your overall thoughts about the questionnaire?
- Is there anything we forgot to ask that is important to patients seeking treatment?
- Is there anything we should change about our questionnaire?
- Is there anything else that you would like to add or comment on?

Part 2: concept elicitation for new scales

Experience of care

- Can you tell me about the first time you asked for professional help with your gender identity?
- Can you tell me about the events that led to your decision to seek treatment?
- Have you experienced any barriers to obtaining gender-affirming treatments?
- What are the people like who have cared for you? Probe: helpful, friendly, expertise, skill.
- What kind of information and advice were you given by healthcare providers? Probe: amount, content, interactions.
- How could healthcare professionals improve the quality of care provided to you?

Treatments and recovery

- What kinds of gender-affirming treatments have you had so far? Probe: psychological, hormonal, surgical.
- What was the recovery like from surgical treatments? Probe: symptoms and impact on activities and daily life.
- What was good/bad about each gender-affirming treatment you have had?
- How happy/satisfied are you with each gender-affirming treatment you have had?

- Do you plan to have any gender-affirming treatments in the future? If yes, which ones?

Appearance

- How important is your appearance to you?
- To what extent is your appearance aligned with your gender identity?
- How would you describe your appearance? Probe: face/body/hair/other.
- What do you like/dislike about the appearance? Probe: face/body/hair/other.
- Has your appearance changed with any gender-affirming treatments and how? Probe: face/body/hair/other.
- Is there anything else about your appearance that you would like to change? Probe: face/body/hair/other.

Voice

- How important is how your voice sounds to you?
- To what extent is your voice aligned with your gender identity?
- How would you describe your voice?
- What do you like/dislike about your voice?
- How has your voice changed with any gender-affirming treatments and how? Probe: do people respond to you differently?

Psychological

- How do you generally feel? Probe: negative (anxiety, depression, irritation) and positive (happy, post-traumatic growth).
- Does distress interfere with daily activities and how? Probe: work, social, dating.
- Has your emotional health changed with any gender-affirming treatments and how?
- Which gender-affirming treatments have helped the most in terms of emotional well-being?

Body image

- How do you feel about your face/body/hair/other? Probe: negative (self-conscious, uncomfortable, unattractive, abnormal) and positive (confident, attractive, normal).
- Do you hide or cover parts of your body or face or modify your voice to blend in? Probe for how and why.
- To what extent is your body image in line with your gender identity?
- Has your body image changed with any gender-affirming treatments and how?
- Which gender-affirming treatments have helped the most in terms of body image?

Social

- What has it been like for you socially? Probe: school, work, friends, family, dating.
- Can you describe the kinds of emotional, informational and instrumental support you receive from your social network?
- Have you experienced any bullying, teasing, abuse, stigmatisation or discrimination? Probe: school, work, friends, family, dating and coping strategies.
- Are there things you want to do in life but do not because of gender-specific issues?
- Has your social life changed with any gender-affirming treatments and how? Probe: school, work, friends, family, dating.
- Which gender-affirming treatments have helped the most in terms of your social life?

Physical

- To what extent is your physical function aligned with your gender identity? Probe: mobility, strength, activities.
- How has your physical function changed with the gender-affirming treatments? Probe: mobility, strength, activities.
- Which gender-affirming treatments have helped the most in terms of physical function?

Sexual

- How important is your sexual life and well-being to you?
- How do you feel about your body sexually? Probe: positive (confident, attractive, normal) and negatives (self-conscious, uncomfortable, unattractive, abnormal) and clothed/unclothed.
- Can you tell me about any problems or concerns with the sexual life related to your gender identity?
- In what ways has gender-affirming treatment changed your genitalia? Probe: appearance, function.
- Has your sexual life changed with any gender-affirming treatment and how? Probe: frequency, quality, drive, arousal.
- Which gender-affirming treatments have helped the most in terms of your sexual life?

Additional questions

- Can you describe any other concerns or issues you experienced that we have not already covered?

Thinking back over this interview, what are the most/least important issues that we have talked about?

Supplementary material 2: GENDER-Q FIELD-TEST SCALES AND THE NUMBER OF EXPERTS AND PATIENTS TO PROVIDE FEEDBACK ON EACH SCALE

Domain	GENDER-Q Scales	Context of use	Expert review	Patient review
HRQL	Body Image	Any treatment	26	14
	Gender Dysphoria	Any treatment	26	14
	Social Acceptance	Any treatment	27	14
	Psychological Distress	Any treatment	28	12
	Psychological Well-Being	Any treatment	28	12
	Treatment Outcome	All surgery	35	9
Sexual	Sexual Well-Being	If sexually active, Genital surgery	26	14
	Orgasm***	If sexually active, Genital surgery	0	14
Urination	Urination	Any genital surgery	31	7
	Urinary Catheter	Any genital surgery	31	8
Gender Practices	Binding Well-Being	If binding	14	8
	Binding: Adverse Effects	If binding	14	14
	Tucking: Adverse Effects	If tucking	23	7
Voice	Sound	Speech surgery or therapy	22	11
	Distress	Speech surgery, therapy	22	11
Hair	Hair – Face	Hair removal or addition	28	7
	Hair – Face [§]	Hair removal or addition	28	7
	Hair - Head	Scalp advance, transplant	27	14
Face & Neck	Face	Female feminization	25	7
	Upper Face	Female feminization	25	7
	Nose	Rhinoplasty	26	7
	Nostrils	Rhinoplasty	25	7
	Lips	Surgery or filler	25	7
	Cheeks**	Implant, fat graft	16	7
	Chin	Surgery, implant	25	7
	Jawline	Surgery, shave	25	7
	Adam's Apple	Tracheal shave	26	7
Body	Body**	Contouring, liposuction	15	7
	Buttocks**	Contouring, liposuction	14	7
	Waist**	Contouring, liposuction	14	7

Breasts	Breasts	Feminizing top surgery, fat grafting	28	7
	Nipples Areolas	Masculinizing top surgery, fat grafting	29	7
	Animation Deformity**	Submuscular implants	20	7
Genital Feminization	Vagina	Vaginoplasty	29	7
	Labia	Labiaplasty	29	7
	Clitoris	Feminizing genital surgery	29	7
Chest	Chest	Masculinizing top surgery	30	7
	Nipples/areolas	Masculinizing top surgery, fat grafting	29	7
	Scars**	Masculinizing top surgery	19	7
Genital Masculinization	Penis	Phalloplasty, Metoidioplasty	31	7
	Penis Sensation	Phalloplasty, Metoidioplasty	32	7
	Glans	Glansplasty	28	7
	Scrotum	Scrotoplasty	29	7
	Perineum	Surgery to create perineum	29	7
	Phalloplasty Flap	Phalloplasty	29	7
	Phalloplasty Scars**	Phalloplasty	29	7
	Donor Site: Adverse Effects	Phalloplasty	29	7
	Testicular implants	Phalloplasty, Metoidioplasty	29	7
	Erectile device	Phalloplasty	29	7
Experience of care	Healthcare Professional	Quality improvement	37	11
	Clinic	Quality improvement	37	10
	Surgery - Information	Quality improvement	35	10
	Surgery - Adverse Effects	Any surgery	31	12
	Surgery - Return to Activity	Any surgery	30	14

**Scales added after round 1 of scale refinement and reviewed by up to 26 experts.

***This scale was only reviewed by patients