Patient-Reported Outcome Measures in Gender-Affirming Surgery
Manraj Kaur, PhD, MSc, Shane Morrison, MD, MS, Andrea Pusic, MD, MHS, and Anne Klassen, DPhil

Abstract
Patient-reported outcome measures (PROMs) are questionnaires that assess how patients feel and function. PROMs should be developed and validated using a mixed methods, multistep approach with extensive patient input to ensure that they are easy to understand, comprehensive, and relevant. PROMs that are specific to gender-affirming care (including surgery), such as the GENDER-Q, can be used to educate patients, align patients’ goals and preferences with realistic expectations about the surgical procedures’ purposes and outcomes, and conduct comparative effectiveness research. PROM data can contribute to evidence-based, shared decision making and just access to gender-affirming surgical care.

The Importance of Asking Patients
Gender-affirming surgery includes a range of individualized and medically necessary procedures that are performed to align an individual’s physical characteristics with their gender identity. Demand for gender-affirming surgery has grown exponentially in recent years, with 25% of transgender and gender diverse (TGD) individuals reporting in a 2015 survey that they had undergone some type of gender-affirming surgery. In parallel, there has been an upsurge in gender-affirming surgical options and technical variations. Gender-affirming surgeries are often complex, as they can involve multiple specialties, and might be irreversible. They are also associated with high costs to the health care system and the patient (eg, copays). Consequently, to provide the highest-quality and evidence-based care, it is crucial to measure and longitudinally evaluate outcomes of gender-affirming procedures and to conduct comparative effectiveness research.

To date, the measurement of outcomes in the gender-affirming surgery literature has largely focused on the clinician perspective (ie, clinical judgment or interpretation of a patient’s observable signs or physical manifestations of a condition). These clinician-reported outcomes are impairment focused and include, for example, wound healing, bleeding, nerve injury, and flap loss. However, only collecting and reporting clinician-reported outcomes overlooks the impact of gender-affirming surgeries and related complications on patients and their health-related quality of life. Patient-reported
outcomes (PROs) are unobservable or latent outcomes known only by the patient and cannot be assessed using clinical observation or physical examination. PRos are symptom and function focused and may include physical symptoms (eg, pain, fatigue), functions (eg, activities of daily living, sleep, work), psychosocial well-being, and sexual well-being. These outcomes are measured using standardized and validated questionnaires (also called scales, surveys, or instruments) without the data being interpreted by a health care professional or anyone else and are called patient-reported outcome measures (PROMs). PROMs, including for gender-affirming care, have a number of benefits and should be developed and validated using a mixed methods, multistep approach with extensive patient input to ensure that they are easy to understand, comprehensive, and relevant.

**PROMs Benefits**

At its core, the use of PROMs allows for systematic and meaningful inclusion of patient voice in treatment decision making and enhances patient-centered care. However, collecting and utilizing PROM data may have a multilevel impact on how health care is planned, organized, delivered, and reimbursed (see Figure).7,8,9

**Figure. Multilevel Uses of PROMs Data**

Previous studies have shown that completing a PROM can result in patients’ improved awareness of their health status or treatment-related effects and provide patients with relevant terminology (nano level), enabling them to better communicate with their health care team.10,11,12 At the level of patients and health care professionals (micro level), PROM data can be used to set expectations or align a treatment approach with the preferences of the patient, educate the patient, facilitate clinician-patient communication, identify pre- or postoperative concerns, prioritize health outcomes, and measure changes in health over time.13 At the level of a health care organization (meso...
level), systematically or routinely collected PROM data can be used to assess health outcomes over time. More specifically, patient data can be used to predict health outcomes for clinical and sociodemographic subgroups and to evaluate the comparative clinical effectiveness of treatment interventions. The PROM data also can be used to evaluate clinician performance and for peer benchmarking.14,15 The organization may use these data to evaluate program effectiveness and efficiency as well as quality assurance and improvement initiatives and to identify gaps in health care services. Lastly, PROM data are useful to health care systems (macro level) in comparing health outcomes across different organizations or jurisdictions for the purpose of informing health care reimbursement and policy decisions, ultimately providing the basis for value-based reimbursement.16,17

PROM Design
Broadly, there are 2 main types of PROMs: (1) generic PROMs that measure overall health or well-being or general aspects of health status and (2) condition- or treatment-specific PROMs that measure symptoms and symptom interference for a specific condition or treatment. Both generic and condition- or treatment-specific PROMs are required to meet PROM development and validation guidelines that have been put forth by the US Food and Drug Administration (FDA),18 COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments),19 the Professional Society for Health Economics and Outcomes Research (ISPOR; formerly, the International Society for Pharmacoeconomics and Outcomes Research),20,21 and similar organizations.22,23 At a minimum, the guidelines recommend that the development of a PROM should begin by defining the construct, target population, and context of use. As part of this process, extensive qualitative input should be sought from the people who experience the construct—and for whom the measure is intended—to establish the PROM’s face validity (what the PROM appears to measure from patients’ perspectives). Additionally, patient data should be used to develop questions (ie, items) for the PROM. The items should include words used by patients as much as possible, and any double-barreled, technical, or value-laden terms should be avoided. Once the items are developed, appropriate response options, recall duration, and instructions should be defined. The PROM should be piloted among patients using cognitive debriefing interviews, and expert feedback should be sought to establish the PROM’s content validity (ie, comprehensibility, comprehensiveness, and relevance). A field test study should be conducted with a large, heterogeneous sample of patients to assess the PROM’s reliability (ie, internal consistency, measurement error), construct or criterion validity (whether the PROM measures what it intended to measure), as well as its responsiveness (whether the PROM captures change over time in health status or condition).19 Scoring algorithms should be established based on the theoretical approach guiding PROM development and validation. Following this process, the PROM should be made available for clinical care and research. The PROM may be translated into other languages and culturally adapted for increased uptake using ISPOR’s best practice guidelines.24

PROM Data Collection and Implementation Considerations
PROM data collection should always start with W5H questions—why, who, what, when, where, and how (see Table). Establishing concordance between what matters to the target population (the construct of interest) and what the PROM is intended to measure is of utmost importance for a successful PROM data collection program. A core team of key stakeholders—patients, clinicians, researchers, payers, regulators, and, where applicable, caregivers, hospital administrators, or community organizations—should be established and their feedback integrated into the planning, design, implementation,
and evaluation of the program. The feasibility (ease of implementation, practicality, integration with information technology such as electronic health records, and scalability) and acceptability (face validity, content validity, ethics, burden, opportunity cost) of the PROM(s) should be examined in a pilot study prior to scaling PROM data collection at the meso and macro levels.25,26

<table>
<thead>
<tr>
<th>Question</th>
<th>Explanation</th>
</tr>
</thead>
</table>
| Why      | Establish a clear purpose for PROM data collection and how the data will be used.  
  • Diagnose, prescribe, predict, or evaluate. |
| What     | Define the construct of interest and identify a suitable PROM.  
  • Ensure that the construct of interest is aligned with what matters to patients who are seeking or receiving gender-affirming surgery.  
  • The PROM should have content and face validity, in addition to established reliability and validity in the target population.  
  • Other considerations when choosing a PROM include available translations and fees to use the PROM. |
| Who      | Define the target population in which data will be collected.  
  • Ensure that all relevant subgroups are accurately represented and that the data collection procedures are accessible and equitable. |
| When     | Establish the timing and frequency of data collection.  
  • These parameters predominantly should be dictated by the purpose of data collection and the clinical research question.  
  • Ensure accessibility and prevent research fatigue. |
| Where    | Collect remotely or at the point of care (eg, prior to a clinic visit).  
  • Ensure accessibility and reasonable privacy for patients completing PROMs. |
| How      | Develop and pilot the plan for data collection.  
  • Data may be collected via paper, electronic (desktop or portal) devices, telephone (eg, automated voice recognition software) or a combination thereof.  
  • Data may be collected before or during clinic visits.  
  • Consider administrative burden, cost, accessibility, and real-time use of PROM data for clinical care.  
  • Consider information technology-related requirements. |

Implementation of PROMs in gender-affirming surgery at the hospital, program, system, or national level should be grounded in implementation science frameworks (deterministic and evaluative) with a focus on intersectionality (eg, the Consolidated Framework for Implementation Research27,28,29 enhanced for intersectionality30). Prior to implementation, extensive input should be sought from all stakeholders on factors that affect implementation success and scalability, including barriers to and enablers of PROM data collection, such as staff and organizational preparedness. The clinic workflows should be refined to ensure minimal logistical burden to clinic staff and patients, and the clinic staff should be trained on the collection, interpretation, and use of PROM data. Information technology-related resources (eg, data reporting, analytics) should be harnessed or developed to ensure accessible and equitable data collection. An iterative feasibility evaluation should be conducted to ensure that the preset quality indicators (eg, program fidelity, PROM completion rates) are met and that there are no
gaps in the efficient and effective scaling of the PROM data collection program. Elements of PROM programs that have been linked to long-term success include identifying clinical champions, dedicated staff members and resources, ensuring stakeholders’ commitment to integrate and use PROM data, accessibility of PROM data for clinical care, and actionable feedback to patients and clinicians based on PROM data. Guidance is available for planning PROM implementation and selecting PROMs implementing and evaluating PROM initiatives, integrating PROMs into electronic health systems, and visualizing PROM data.

**Patient-Facing Policy**

A key consideration in gender-affirming surgery is that PROs research, as it expands, should aim to reduce health- and health care-related disparities at the policy level. Efforts should be taken at the micro, meso, and macro levels to ensure that PROMs are designed and implemented in fully accessible ways and without the unintended exclusion or inundation of patient subgroups. PROMs should be made available in languages spoken by patients, require no more than a sixth-grade reading level, and employ hybrid modes and methods of data collection (eg, during a clinic visit or remotely, on mobile devices or on paper). The environment in which PROMs are administered or used should foster inclusiveness by ensuring that the staff are culturally competent, by providing accessible spaces (eg, gender-neutral washrooms), and by using intake forms that include a variety of gender and sexual identities. Data collected at the hospital system (meso) or jurisdiction (macro) level should be analyzed to identify ways to improve care quality and cost effectiveness to promote value-based health care. The analysis, use, and dissemination of PROM data at all levels should thoroughly and thoughtfully consider the impact of the data on extant health policies that fund and regulate access to gender-affirming care.

**PROMs in Gender-Affirming Surgery**

Although PROMs have been used to assess gender-affirming surgery outcomes for the last few decades, recently the shortcomings related to the development and psychometric properties of existing PROMs have been called to attention. Converging evidence from recent systematic reviews on PROMs used in gender-affirming surgery highlight 4 key issues. First, most PROMs identified in the literature were developed to be used for a specific study and therefore lack validation. Second, several PROMs that are used in the gender-affirming care literature were developed to evaluate outcomes in cisgender groups and have not been rigorously validated in gender diverse individuals (eg, the Female Genital Self-Image Scale, the International Prostate Symptom Score). Third, the number of PROMs used in the gender-affirming surgery literature are limited by their content or by failing to follow international guidelines for PROM development. Lastly, PROMs that comprehensively assess outcomes of specific types of treatment interventions or procedures (eg, scrotoplasty, labiaplasty) or a single body part or region (eg, forehead, jaw, facial hair) are lacking. An urgent need for a comprehensive, rigorously designed, and validated PROM to assess outcomes of gender-affirming care was identified. Our international team of clinicians, quality-of-life researchers, and psychometricians responded to this call to action by developing the GENDER-Q—a PROM for assessing outcomes in gender-affirming care.

The GENDER-Q consists of a comprehensive set of unidimensional scales (questionnaires) that assess the domains of appearance (hair, face, neck, body, breasts, chest, genitals, donor site), health-related quality of life (physical, psychosocial, sexual, voice, practices), the experience of care (health professional, clinic, preoperative
information, and outcome), and devices (catheter, testicular implants, erectile devices). To develop the GENDER-Q, our team followed international guidelines for PROM development. We conducted in-depth interviews with 84 TGD individuals from 4 countries (Canada, Denmark, the Netherlands, and the United States) who were seeking or had undergone gender-affirming treatment(s). The data were used to create the preliminary versions of a set of independently functioning scales. The scales were shown to 7 to 14 TGD individuals (depending on the scale) and 50 clinicians and research experts and iteratively refined, resulting in the field test version of the scales. An international field test study to establish the measurement properties of the GENDER-Q is underway. The data collected will be used to refine the scales, assess their reliability and validity, and develop a common scoring algorithm for each scale for international use. Once the field test is completed, the scales and scoring will be made available for not-for-profit clinical research and care at no charge.

The GENDER-Q represents a positive ethical shift in the measurement of PROs for gender-affirming surgery, as it lays the foundation for a patient-centered health care culture that promotes the notion of “nothing about us without us,” as opposed to the current, fundamentally flawed practice of using PROMs in gender-affirming surgery that were developed for the cisgender population.

Conclusion
Empirically and systematically integrating rigorously developed and validated gender-affirming surgery-specific PROMs (eg, GENDER-Q) that capture what matters to patients are indispensable to patient-centered, shared treatment decision making; improving care quality; and expanding access to and funding of surgical procedures.

References


**Manraj Kaur, PhD, MSc** is a Canadian Institutes of Health Research-funded fellow at the Patient-Reported Outcomes, Value, and Experience Center at Brigham and Women’s Hospital in Boston, Massachusetts. Dr Kaur has experience in the measurement and implementation of patient-reported outcomes in clinical research and practice and in conducting micro-costing and cost-effectiveness studies.

**Shane Morrison, MD, MS** is an assistant professor of plastic and gender affirmation surgery at Seattle Children’s Hospital and the University of Washington Medical Center. Dr Morrison’s clinical practice spans adolescent and adult gender affirmation, and his research is focused on surgical and patient-reported outcomes.

**Andrea Pusic, MD, MHS** is the chief of the Division of Plastic and Reconstructive Surgery and the director of the Patient-Reported Outcomes, Value, and Experience Center at Brigham and Women’s Hospital in Boston, Massachusetts. Dr Pusic’s research involves the development and integration of patient-reported outcome measures for use in pediatric and adult patients undergoing plastic and reconstructive surgery, as well as quality improvement and comparative effectiveness research.

**Anne Klassen, DPhil** is a professor in the Department of Pediatrics at McMaster University in Hamilton, Ontario, Canada, and an associate member of the Department of Surgery and the Department of Health Research Methods, Evidence, and Impact. Dr Klassen has worked on the development of the Q-Portfolio series of patient-reported outcome instruments for pediatric and adult patients undergoing plastic and reconstructive surgery.
Citation

DOI

Conflict of Interest Disclosure
GENDER-Q was co-developed by Drs Anne Klassen, Manraj Kaur, and Andrea Pusic; Brigham and Women’s Hospital and McMaster University own copyright. Dr Klassen could receive license revenue when GENDER-Q is used in investigations. Dr Pusic co-developed Q-PROM portfolio measures and could receive royalties when they are used in clinical trials. Dr Morrison had no conflicts of interest to disclose.

The viewpoints expressed in this article are those of the author(s) and do not necessarily reflect the views and policies of the AMA.