Published in final edited form as:

Plast Reconstr Surg. 2019 March ; 143(3): 679–684. doi:10.1097/PRS.0000000000005314.

Introducing BREAST-Q Computerized Adaptive Testing – short and individualized patient-reported outcome assessment following reconstructive breast surgery

Danny A. Young Afat, MD¹, Christopher Gibbons, PhD^{2,3}, Anne F. Klassen, DPhil⁴, Andrew J. Vickers, PhD⁵, Stefan J. Cano, PhD⁶, Andrea L. Pusic, MD^{2,3}

¹University Medical Center Utrecht, the Netherlands ²Brigham and Women's Hospital, Boston, USA ³Harvard University, USA ⁴McMaster University, Canada ⁵Memorial Sloan Kettering Cancer Center, USA ⁶Modus Outcomes, UK

Abstract

BACKGROUND—BREAST-Q is a widely used patient-reported outcome (PRO) instrument measuring health-related quality-of-life and patient satisfaction in breast surgery. Shorter assessment potentially increases patients' willingness to complete scales, but simply offering a short-form version leads to unacceptable loss in measurement precision. We aimed to develop a computerized adaptive test (CAT) to shorten BREAST-Q's Satisfaction with Breasts scale whilst maintaining reliability of measurement.

METHODS—We created a CAT, which repetitively administered questions from the pool of 16 questions, until pre-specified levels of reliability were reached (i.e., standard errors (SE) of 0.32–0.55). In a simulation study, we tested the CAT's feasibility for all potential Satisfaction scores. In a second study using actual patient data, 5000 breast reconstruction patients who had previously completed the full scale were randomly selected from a large database. Their full-scale Satisfaction scores were compared with their CAT-derived scores.

RESULTS—In both studies, by applying CAT, the Satisfaction with Breasts scale could be reduced to an average of 10 questions when using the minimum level of measurement precision for individual-patient measurement (SE 0.32), compared to 4 questions when using the minimum precision level for group-based research (SE 0.55). Score estimates were highly correlated

Corresponding author: Christopher Gibbons, Patient-Reported Outcomes, Value, and Experience (PROVE) Center, Department of Surgery, Brigham and Women's Hospital, 75 Francis Street Boston, MA 02115, Fax 617-732-6387, cgibbons2@bwh.harvard.edu.

Meetings:

Preliminary data was presented at the 23rd Annual Conference of the International Society for Quality of Life (ISOQOL) Research meeting in Denmark in 2016, the Patient Reported Outcome Measures UK conference in the United Kingdom in 2016, and at the bi-annual meeting of the Dutch Society for Plastic Surgery (NVPC) in the Netherlands in 2016.

Disclosures:

Andrea Pusic, Anne Klassen and Stefan Cano are co-developers of the BREAST-Q, which is owned by Memorial Sloan-Kettering Cancer Center and the University of British Columbia. They receive a portion of licensing fees (royalty payments) when the BREAST-Q is used in industry sponsored clinical trials. Stefan Cano is co-founder of Modus Outcomes, an outcomes research and consulting firm that provides services to pharmaceutical, medical device, and biotechnology companies. The remaining authors have no financial disclosures.

between CAT assessment and the full scale (0.91–0.98 in simulation study, 0.89–0.98 in patient data study).

CONCLUSIONS—Applying CAT to BREAST-Q’s Satisfaction with Breasts scale facilitates reliable assessment with 38% to 75% fewer question than the full version. The novel BREAST-Q CAT version may decrease response burden and help overcome barriers to implementation in routine care.

INTRODUCTION

Breast reconstruction rates have rapidly increased in the past decade, and the cosmetic outcome is now a common subject when discussing surgical options for breast cancer [1,2]. Patient-reported outcomes (PROs) have become essential to determine the success of breast reconstructions since their primary surgical goal is to improve body image and health-related quality of life (HR-QoL) [3,4]. PROs are usually assessed using validated questionnaires, designed to quantify HR-QoL and other important outcomes from the patient’s perspective [3]. PROs are used for research (e.g., which treatment works best) and for clinical care (e.g., how is this patient doing and should we change the course of treatment?).

When collected in routine clinical care, PROs may help clinicians and patients gain timely insight into patients’ symptoms, preferences and progress; a process shown to improve patients’ self-efficacy, HR-QoL, the patient-physician dialogue and even survival [5–11]. Sharing PRO scores from patients with similar preferences may help guide their decisions (e.g. breast-conserving surgery versus mastectomy with reconstruction), while evaluating PROs in specific subgroups could improve patient selection for the different surgical options (e.g., implant reconstruction versus autologous reconstructions in the setting of radiotherapy) [2].

In order to use PROs for these important purposes, the PRO-instruments need to be scientifically sound, able to detect changes over time, and should measure outcomes that matter to patients in the area of interest. Generic instruments, and even cancer-specific questionnaires, may thus not be sensitive enough to evaluate relevant surgical outcomes (e.g., cosmetic satisfaction, complications specific to the surgical area) [3,4].

With these limitations in mind, the BREAST-Q was developed to measure HR-QoL and patient satisfaction for patients undergoing breast surgery [13]. The BREAST-Q has shown to be valid, psychometrically robust and capable of detecting clinically meaningful changes over time [12–15]. Since its inception in 2009, the BREAST-Q has been widely used in surgical research and clinical trials, clinical practice, and quality improvement initiatives [2,15–17]. The BREAST-Q scale that measures patient satisfaction with breasts has become part of the ICHOM Standard Set for Breast cancer [4].

The BREAST-Q consists of scales that can be used independently. Each scale is scored individually, and there is no total score for all scales combined. This allows investigators and clinicians to use only those scales that are relevant in a given study or clinical scenario (e.g., ‘Physical well-being’ during follow-up visits, ‘Satisfaction with Breasts’ after scars have healed). Despite this system of independently functioning scales, in certain clinical and

research situations, there may be a need for even shorter assessments (e.g., to reduce the patient burden when many different questionnaires are being administered, or to facilitate quick assessment in a busy clinical setting). The need to reduce response burden may especially apply to oncology, where patients are often overwhelmed with complex information to process and many forms to complete.

Simply offering a ‘short-form version’, however, may lead to unacceptable loss of precision and an inability to measure smaller clinically relevant changes over time [18,19].

Computerized adaptive testing (CAT) is an innovative solution used to individualize and shorten (existing) questionnaires, whilst maintaining reliable estimates for scores [18–29]. Instead of presenting all questions in a fixed order, CAT uses an algorithm to decide which question will be presented to the patient next. Based on each answer, the most informative and most appropriate question is then selected until a pre-specified level of measurement precision is reached. For some patients, this level could be reached after a few questions, while other patients need to complete all questions in a scale. As a result, CAT may not only shorten scale completion time, but may also prevent the presentation of questions that are less relevant to an individual patient [18–30].

We aimed to develop a CAT for the BREAST-Q ‘Satisfaction with Breasts’ scale of the reconstruction module, and to assess its performance in terms of efficiency, precision, and comparability to the full scale.

METHODS

BREAST-Q and Rasch Measurement Theory

BREAST-Q was developed according to international guidelines for PRO instrument development [7,12,31,32], and its domains and questions were constructed with extensive input from patients. Rasch Measurement Theory (RMT) was used to guide scale development and refinement. In general, scales that fit the Rasch model have an ordering of questions in the same way that numbers increment along a ruler, with an equal distance between each increment (i.e., interval scaling). Furthermore, the ordering of questions follows a clinical hierarchy. Questions that correlate with, for instance, lowest HR-QoL are on the far left of the ruler, while questions that correlate with highest HR-QoL are located on the far right of the ruler. For example, only patients with the most symmetrical autologous reconstructions will likely confirm being satisfied or very satisfied when looking in the mirror without their clothes on (the question on the far right of the Satisfaction with Breasts ‘ruler’).

This approach to scale development has become popular in the past decade, as it enhances the ability to understand the clinical meaning of a patient’s score, since the score for each scale can be interpreted as a location on the ‘Rasch ruler’. This facilitates the interpretation of findings both on a group level (i.e., relevant for research) and on an individual level (i.e., relevant for measurement and follow-up in clinical care).

In comparison, PRO instruments developed using the more traditional Classical Testing Theory – which applies to the majority of PRO instruments available for breast cancer – are

mainly valid for use in group- or population-based research [21,22,25,27]. Also, in the Classical Testing approach, statistical analyses are conducted at the scale level rather than question level, meaning that all questions must be delivered to all patients, often resulting in long scales. In contrast, Rasch analysis produces precise psychometric information about individual questions within a scale leading to greater flexibility in the manner in which the scale is administered, including the potential to administer scales using computerized adaptive testing. A more in-depth description of the Rasch model, Classical Testing Theory and psychometrics can be found in other sources [21,22,25,27,31].

Computerized adaptive testing

Computerized adaptive testing may be applied to scales that have been developed using Rasch analysis – such as the BREAST-Q. Each question is calibrated to a position on the Rasch ruler. As a result, the answer to each question helps determine a patient's position on that ruler. Instead of presenting all questions in a fixed order, CAT uses an algorithm to decide which question will be presented to the patient next. After each answered question, the CAT uses the current position on the ruler (i.e., Rasch-derived item location) to determine a provisional score. The CAT then selects the question that will provide the most additional information about this provisional score. Only when another question provides relevant additional information would that question be presented next. As a result, the assessment is tailored to each individual patient [18–25]. The CAT algorithm continues to administer questions until a pre-determined level of measurement precision (known as the stopping rule) has been reached. Once the stopping rule has been reached, the assessment is completed and the final score is calculated.

The BREAST-Q CAT

For this study, we tested CAT feasibility for the 'Satisfaction with Breasts' scale, as this is one of the longer and most widely used BREAST-Q scales of the reconstruction module [12,13,15]. The fixed-order full scale consists of 16 questions. The question on the far left of the Rasch ruler evaluates 'How you look in the mirror clothed', while the question on the far right evaluates 'How you look in the mirror unclothed'.

Psychometric properties (e.g., where each question's response options are located along the ruler) were used to create a CAT for this scale. These properties were also used to determine the opening question of the CAT (usually a question that is located around the middle of a scale's Rasch ruler or around the scale's mean population score) [19, 25–28].

After a patient responded to the opening question, the CAT algorithm first estimated the patient's provisional Satisfaction with Breasts score. This provisional Satisfaction score was calculated with a surrounding confidence interval, known as the standard error (SE). After each next question was answered, a new provisional Satisfaction score and SE were calculated. As each response added more information on how satisfied the patient was with their reconstructed breast(s), the SE of the score estimate decreased. The CAT continued to administer questions until the SE fell below the pre-specified stopping rule, after which the assessment ceased, and a final score was calculated.

To assess the performance of the CAT in a variety of situations, stopping rules were set to 3 different standard errors (i.e., 0.32, 0.45 and 0.55). These SE values were chosen because they are equivalent to reliability values (i.e., Cronbach alpha) of 0.90, 0.80 and 0.70 respectively (reliability = $1 - SE^2$). A reliability of 0.70 is generally seen as the minimum value for reliable group measurement, whereas the minimum for individual-level measurement is 0.90 [19,20,28].

We performed two studies with the CATs. During the first study, all possible values for the Satisfaction score on the Rasch ruler were simulated with intervals of 0.10. The CATs were created using R coding, and the simulations were conducted in R Studio using modified FireStar code. During the CAT process, each question was only administered once.

To compare scores retrieved by CAT with scores from the fixed-order full scale, we conducted a second study using actual patient data. Five thousand patients were randomly selected from a larger sample of more than 17,000 patients who had previously completed the 'Satisfaction with Breasts' scale. These data had previously been collected prospectively in an industry-sponsored post-marketing breast implant surveillance study. These women had all undergone implant-based breast reconstructions in the United States after a breast cancer diagnosis, were 22 years or older, and fluent in English. For these 5000 randomly selected patients, 'Satisfaction with Breasts' scores were initially calculated based on their full-scale responses and separately by using their responses to run the CAT. As a result, this second study emulates what would have happened if the CAT were used when the patient was completing the scale. Pearson correlation coefficient was calculated to assess comparability between scores.

RESULTS

Psychometric properties retrieved by Rasch analysis of the original BREAST-Q field test data confirmed suitability for CAT. More details about these properties have been described elsewhere [12–14]. The question that provided the most information at the population mean for the 'Satisfaction with Breasts' scale of BREAST-Q's Reconstruction module was 'How equal in size your breasts are to each other?'. Therefore, this question was chosen as the opening question of the CAT for all patients.

With CAT, the 'Satisfaction with Breasts' scale could be reduced to an average of 10 questions when the minimum level of measurement precision for individual-level measurement was chosen (SE 0.32). This provided a reduction of 37.5% in questions as compared to the full 16-item scale, with a correlation of 0.98. The scale was reduced to an average of 4 questions when using the minimum acceptable level for group-based research (SE 0.55), providing a reduction of 75%, with a correlation between the final scores of 0.91.

In a second step, the CATs were applied to the existing full-scale responses of the random sample of breast reconstruction patients (n=5000) as if their responses had been collected adaptively. These results were equal to the simulation study, showing that CAT-derived Satisfaction scores from actual patients were also only marginally different from the scores based on full-scale data (Pearson's correlation 0.89–0.97). Results from both studies are

summarized Table 1. Details of individual question parameters used in these simulations are available from the corresponding author.

DISCUSSION

This study shows that applying CAT to BREAST-Q's Satisfaction with Breasts scale of the reconstruction module shortens the scale substantially (38–75%) whilst maintaining reliable outcomes.

Although the reduction in the number of questions is appealing, we do not recommend using BREAST-Q CAT for every situation. There will still be situations where the highest level of measurement precision is required. For example, when robust PRO measurement is to be obtained, such as in certain Randomized Controlled Trials or perhaps also to inform certain value-based healthcare systems. There may also be clinical scenarios where clinicians will be interested in seeing responses to particular questions that the CAT may decide to skip. It is also important to note that BREAST-Q CAT can only be applied when the scale is being administered on a computer (or smartphone/tablet).

In recent years there has been a proliferation of studies using CAT to improve the efficiency of PRO instruments that were also developed using similar psychometrics methods (i.e., RMT, Item Response Theory). The global Patient-Reported Outcome Measurement Information System (PROMIS) initiative has produced a large number of item banks to measure broad aspects of patient-reported health and QoL. Our results concur with their results, and those of other CAT investigations, which all demonstrate the effectiveness of CAT as a tool to reduce the length of PRO assessments whilst maintaining reliability of measurement. Overall, CAT assessments appear to take less time than fixed-form full versions of the same PRO-instruments [18–30].

An important strength of this study is that we performed two different kinds of studies to evaluate the CAT - a first simulation to test CAT feasibility for all potential Satisfaction with Breasts scores, and a second study to test the CAT using actual patient data. Our findings gave us insight into the expected reduction in questions for the CAT in a real-life situation. Both studies gave similar results.

An important limitation of this study is that we simulated CAT responses derived from a previously collected sample. Therefore, we were unable to collect information on completion time for the BREAST-Q CAT in a true administration environment. Such Information could be relevant to determine whether fewer questions in a scale will also reduce *actual* completion time. As the questions of the fixed-form BREAST-Q and the CAT are presented in the same format, we expect the percentage of time saved during CAT administration to be roughly equal to the scale reduction percentage due to CAT [33].

There is a trade-off between assessment length and precision. Although the ratio between the two is optimized using the CAT algorithm, decisions regarding stopping rules are contextual and thus should be decided upon by researchers/clinicians. Scores from the fixed-form full version and the CAT version of the Satisfaction with Breasts scale are directly comparable. As such, we recommend that investigators who are new to PRO assessment and/or

BREAST-Q first orientate themselves using the more straightforward fixed-form version of the BREAST-Q. Once familiar, we recommend progression to the BREAST-Q CAT to maximize the efficiency of each assessment. Information needed to create CAT versions of the Satisfaction with Breast scale is available from the corresponding author.

This study was performed using only one BREAST-Q scale from the reconstruction module. Psychometric properties from the other scales confirm suitability for CAT. However, CAT adds less benefit to shorter scales, due to the limited information available from a smaller pool of questions. Adaptive assessments may therefore be less relevant for some of the other BREAST-Q scales (e.g. 10-item psychosocial well-being scale, and the 6-item sexual well-being), which are already quite brief.

Recent studies have demonstrated the clinical benefit – even an association with longer survival – of symptom monitoring using PROs during routine cancer treatment [9–11]. The enhanced efficiency of the BREAST-Q CAT could serve to further break down barriers to using and implementing PROs in breast surgery, potentially improving clinical care and providing answers to important questions that can only be answered by patients themselves [2,16,17].

Acknowledgements:

Andrea Pusic, Anne Klassen and Stefan Cano are co-developers of the BREAST-Q, which is owned by Memorial Sloan-Kettering Cancer Center and the University of British Columbia. They receive a portion of licensing fees (royalty payments) when the BREAST-Q is used in industry sponsored clinical trials. Stefan Cano is co-founder of Modus Outcomes, an outcomes research and consulting firm that provides services to pharmaceutical, medical device, and biotechnology companies. The remaining authors have no financial disclosures.

REFERENCES

1. Rowland JH, Desmond KA, Meyerowitz BE, et al. Role of breast reconstructive surgery in physical and emotional outcomes among breast cancer survivors. *J Natl Cancer Inst*, 2000
2. Jagsi R, Momoh AO, Qi J, et al. Impact of Radiotherapy on Complications and Patient-Reported Outcomes After Breast Reconstruction. *J Natl Cancer Inst*, 2018, 10.1093/jnci/djx148
3. Pusic AL, Chen CM, Cano S, et al. Measuring Quality of Life in Cosmetic and Reconstructive Breast Surgery: A Systematic Review of Patient-Reported Outcomes Instruments. *Plast Reconstr Surg*, 2007 9 15. doi: 10.1097/01.prs.0000278162.82906.81
4. Ong WL, Schouwenburg MG, van Bommel ACM, et al. A Standard Set of Value-Based Patient-Centered Outcomes for Breast Cancer: The International Consortium for Health Outcomes Measurement (ICHOM) Initiative. *JAMA Oncol*, 2017. doi: 10.1001/jamaoncol.2016.4851.
5. Fallowfield L, Jenkins V. Psychosocial/Survivorship Issues in Breast Cancer: Are We Doing Better? *J Natl Cancer Inst*, 2015. doi: 10.1093/jnci/dju335.
6. Runowicz CD, Leach CR, Henry NL, et al. American Cancer Society/American Society of Clinical Oncology Breast Cancer Survivorship Care Guideline. *CA Cancer J Clin*, 2016. doi: 10.3322/caac.21319.
7. US Food and Drug Administration. Health and Quality of Life Outcomes. 2006 Draft Guidance for Industry on Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims URL: <http://www.fda.gov/downloads/Drugs/.../Guidances/UCM193282.pdf> [accessed 2017-08-31]
8. Basch E Patient-reported outcomes — harnessing patients' voices to improve clinical care. *N Engl J Med*, 2017. doi: 10.1056/NEJMp1611252
9. Rotenstein LS, Huckman RS, Wagle NW. Making Patients and Doctors Happier - The Potential of Patient-Reported Outcomes. *N Engl J Med*, 2017. doi: 10.1056/NEJMp1707537

10. Basch E, Deal AM, Dueck AC, et al. Overall Survival Results of a Trial Assessing Patient-Reported Outcomes for Symptom Monitoring During Routine Cancer Treatment. *JAMA*, 2017. doi: 10.1001/jama.2017.7156.
11. Basch E, Dueck AC, Rogak LJ, et al. Feasibility Assessment of Patient Reporting of Symptomatic Adverse Events in Multicenter Cancer Clinical Trials. *JAMA Oncol*, 2017. doi: 10.1001/jamaoncol.2016.6749.
12. Cano SJ, Klassen AF, Scott AM, Pusic AL. A closer look at the BREAST-Q. *Clin Plastic Surg*, 2013 10.1016/j.cps.2012.12.002
13. Pusic AL, Klassen AF, Scott AM, Klok JA, Cordeiro PG, Cano SJ. Development of a new patient-reported outcome measure for breast surgery: the BREAST-Q. *Plast Reconstr Surg*, 2009. doi: 10.1097/PRS.0b013e3181aee807.
14. Cano SJ, Klassen AF, Scott AM, Cordeiro PG, Pusic AL. The BREAST-Q: further validation in independent clinical samples. *Plast Reconstr Surg*, 2012. doi: 10.1097/PRS.0b013e31823aec6b.
15. Cohen WA, Mundy LR, Ballard TN. The BREAST-Q in surgical research: A review of the literature 2009–2015. *J Plast Reconstr Aesthet Surg*, 2016. doi: 10.1016/j.bjps.2015.11.013.
16. Pusic AL, Matros E, Fine N, et al. Patient-Reported Outcomes 1 Year After Immediate Breast Reconstruction: Results of the Mastectomy Reconstruction Outcomes Consortium Study. *J Clin Oncol*, 2017. doi: 10.1200/JCO.2016.69.9561.
17. Dikmans RE, Negenborn VL, Bouman MB, et al. Two-stage implant-based breast reconstruction compared with immediate one-stage implant-based breast reconstruction augmented with an acellular dermal matrix: an open-label, phase 4, multicentre, randomised, controlled trial. *Lancet Oncol*, 2017. doi: 10.1016/S1470-2045(16)30668-4.
18. Jette AM, Haley SM, Pengsheng N, Olarsch S, Moed R. Creating a computer adaptive test version of the late-life function and disability instrument. *J Gerontol A Biol Sci Med Sci*, 2008.
19. Fliege H, Becker J, Walter OB, Bjorner JB, Klapp BF, Rose M. Development of a computer-adaptive test for depression (D-CAT). *Qual Life Res*, 2005. doi: 10.1007/s11136-005-6651-9
20. Gibbons C, Bower P, Lovell K, Valderas J, Skevington S. Electronic quality of life assessment using computer-adaptive testing. *JMIR*, 2016. doi: 10.2196/jmir.6053
21. Cook KF, O'Malley KJ, Roddey TS. Dynamic assessment of health outcomes: time to let the CAT out of the bag? *Health Serv Res*, 2005. doi: 10.1111/j.1475-6773.2005.00446.x
22. Morgan JH, Kallen MA, Okike K, Lee OC, Vrahas MS. PROMIS Physical Function Computer Adaptive Test Compared With Other Upper Extremity Outcome Measures in the Evaluation of Proximal Humerus Fractures in Patients Older Than 60 Years. *J Orthop Trauma*, 2015. doi: 10.1097/BOT.0000000000000280
23. Cella D, Riley W, Stone A, et al. The patient-reported outcome measurement information system (PROMIS) developed and tested its first wave of adult self-reported health outcome items banks: 2005–2008. *J Clin Epidemiol*, 2010. doi: 10.1016/j.jclinepi.2010.04.011
24. Ware JE jr., Kosinski M, Bjorner JB, et al. Applications of computerized adaptive testing (CAT) to the assessment of headache impact. *Qual Life Res*, 2003.
25. Velozo CA, Wang Y, Lehman L, Wang JH. Utilizing Rasch measurement models to develop a computer adaptive self-report of walking, climbing, and running. *Disabil Rehabil*, 2008. doi: 10.1080/09638280701617317.
26. Smits N, Zitman FG, Cuijpers P, den Hollander-Gijsman ME, Carlier IV. A proof of principle for using adaptive testing in Routine Outcome Monitoring: the efficiency of the Mood and Anxiety Symptoms Questionnaire -Anhedonic Depression CAT. *BMC Med Res Methodol*, 2012. doi: 10.1186/1471-2288-12-4.
27. Smits N, Cuijpers P, van Straten A. Applying computerized adaptive testing to the CES-D scale: a simulation study. *Psychiatry Res*, 2011. doi: 10.1016/j.psychres.2010.12.001
28. Flens G, Smits N, Carlier I, van Hemert AM, de Beurs E. Simulating computer adaptive testing with the mood and anxiety symptom questionnaire. *Psychological Assessment*, 2015 10.1037/pas0000240.
29. Hart DL, Mioduski JE, Stratford PW. Simulated computerized adaptive tests for measuring functional status were efficient with good discriminant validity in patients with hip, knee, or foot/ankle impairments. *J Clin Epidemiol*, 2005. doi: 10.1016/j.jclinepi.2004.12.004

30. Cella D, Yount S, Rothrock N, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. *Med Care*, 2007. doi: 10.1097/01.mlr.0000258615.42478.55
31. Patrick DL, Burke LB, Gwaltney CJ, et al. 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*, 2011. doi: 10.1016/j.jval.2011.06.014.
32. Patrick DL, Burke LB, Gwaltney CJ, et al. Content validity - establishing and reporting evidence in newly developed PRO instruments for medical product evaluation: ISPOR PRO good research practices task force report: Part 2—assessing respondent understanding. *Value Health*, 2011. doi: 10.1016/j.jval.2011.06.013.
33. Gibbons CJ. Assessing the impact of computer-adaptive administration, tailored feedback, and response prediction on the user engagement and perceived acceptability of quality of life assessment. *Qual Life Res*, 2017 10.1007/s11136-017-1658-6

Table 1

Summary of CAT studies for BREAST-Q 'Satisfaction with Breasts' scale

	Stopping rule (SE theta)	Number of questions used (mean)	Reliability (Cronbach's alpha)	Correlation full scale score with CAT score
Study 1				
	0.55	4	0.7	0.91
	0.45	7	0.8	0.94
	0.32	10	0.9	0.98
Study 2				
	0.55	4	0.7	0.89
	0.45	7	0.8	0.93
	0.32	10	0.9	0.98

Study 1: All possible values for the 'Satisfaction with Breasts' score were simulated with intervals of 0.10.

Study 2: Comparison between full-scale responses from 5000 actual patients vs. CAT-derived scores when scores were treated as if they had been collected adaptively

NB: full scale consists of 16 questions