Giving meaning to patient reported outcomes in breast reconstruction after mastectomy − A systematic review of available scores and suggestions for further research

Linn Weick [a], Fredrik Brorson [a], Christian Jepsen [a], Mattias Lidén [a],
Emmelie Widmark Jensen [a], Emma Hansson [a, b, *]

[a] *Department of Plastic Surgery, Institute of Clinical Sciences, The Sahlgrenska Academy, University of Gothenburg, Gröna Stråket 8, SE-413 45, Gothenburg, Sweden*
[b] *Department of Plastic Surgery, Sahlgrenska University Hospital, Gröna Stråket 8, SE-413 45, Gothenburg, Sweden*

ARTICLE INFO

ABSTRACT

Background: There are three patient reported outcome measure instruments (PROMs) that have adequate content validity for breast reconstruction, BREAST-Q, BRECON-31 and EORTC QLQ-BRECON-23, and they all have been robustly validated. The aim of this study was to systematically review scores giving meaning to validated PROMs for breast reconstruction after mastectomy and discuss methods to enable interpretation of them.

Methods: A systematic review was performed according to the recommendations of PRISMA. Prospero CRD42021255874. Included articles had to meet criteria defined in a SPIDER (Sample, Phenomenon of Interest, Design, Evaluation, Research type). The included studies were critically appraised using the GRADE approach.

Results: Three articles were finally included in the review: two studies on scores for healthy controls and one on minimally important differences (MIDs), both of BREAST-Q. All of the studies were performed in North America. Only MIDs based on statistical characteristics, and not on what constitutes a relevant change for the patient, exist. The risk of bias was evaluated as very high and moderate, respectively, of inconsistencies as low, of indirectness as high, of imprecisions as low, and of publication bias as probably low.

Conclusions: The overall certainty of evidence for scores giving meaning to PROMs for breast reconstruction is low (GRADE ⊕⊕OO). More studies are needed to establish relevant healthy control scores and what constitutes a relevant clinical difference for patient-reported outcome measures for breast reconstruction after mastectomy. Clinical implications of the findings and suggestions for further research are suggested in the article.

1. Introduction

Patient reported outcomes (PROMs) and quality of life measurements (QoL) have become an essential part of evaluating the result of breast reconstruction [1,2] and are of particular relevance when evaluating interventions specifically performed to improve function, aesthetics, and general well-being [3]. Davies et al. [4] recently performed a systematic review of studies developing and evaluating measurement properties of PROM instruments validated specifically for breast reconstruction after mastectomy. They concluded that there currently are three PROM instruments that have adequate content validity for breast reconstruction, BREAST-Q, BRECON-31, and EORTC QLQ-BRECON-23, and that they all have been robustly validated [4]. Notwithstanding, more aspects need to be studied to make the measurements clinically interpretable and to enable use of them as primary end points in clinical trials [5,6]. For example, in the case of breast reconstruction, what score indicates a relevant clinical difference between different methods, before and after reconstruction, or over time? Moreover, what level of score indicates a reasonable patient satisfaction and a

* Corresponding author. Department of Plastic Surgery, Sahlgrenska University Hospital, Gröna Stråket 8, SE-413 45, Gothenburg, Sweden.
E-mail address: emma.hansson.2@gu.se (E. Hansson).

good result, and are there critical scores (threshold values) suggesting that it is appropriate to perform corrections [5]?

To answer these questions, we need to understand a few theoretical concepts relating to PROMs. The most common ways to enable interpretation of PROM scores are to compare the scores with those of a healthy population ('normative values') and to calculate the smallest change in score that a patient considers important or that makes a clinician modify the treatment [7] ('minimal important differences', MIDs). Comparison with a healthy population requires that scores for that particular PROM are known in a population that is similar to the studied population, both demographically, culturally, as well as health-wise. Ideally, minimal important differences are calculated based on other meaningful changes or 'anchors' ('anchor-based MIDs'). For example, comparison to level of function, symptoms, disease severity, response to treatment, job loss, health care consumption, mortality, impact on life events, global ratings of change, from the patients' or clinicians' perspective, and prognosis of future events (those who experience a certain event versus those who do not) [5,8]. However, most often MIDs are calculated based on different statistical characteristics of the group as an indication of the lowest change value beyond random measurement error ('distribution-based MIDs') [5,8]. Statistically based MIDs vary with characteristics and size of the sample and does not provide any information about the clinical relevance of the change. Often, a combination of both types of calculations are necessary to establish clinically relevant MIDs, so that they are both relevant and outside the measurement error of the instrument [5,8].

The question whether we can interpret the result of PROMs in a meaningful way is a pressing issue due to the widespread use of questionnaires and the considerable effort we ask our patients to make to fill them out. If we ask our patients' time and labour, we must make sure that the responses really add value to the quality of care.

The aim of this study was to perform a systematic scoping review [9] of scores giving meaning to validated PROMs for breast reconstruction after mastectomy, BREAST-Q, BRECON-31 and EORTC-QLQ-BRECON-23: minimal important differences (MIDs) and normative values. Based on the findings, practical and research recommendations going forward will be discussed.

2. Methods

2.1. Protocol

The study protocol was registered in Prospero (CRD42021255874).

2.2. Eligibility criteria for systematic review

Inclusion criteria were studies examining minimal important differences or normative data for BREAST-Q, BRECON-31 and EORTC-QLQ-BRECON-23. The instrument chosen was based on the systematic review performed by Davies et al. [4]. Included articles had to meet criteria defined in a SPIDER (Sample, Phenomenon of Interest, Design, Evaluation, Research type) [10]. Sample: Women who have had a mastectomy and a breast reconstruction or normative populations of women. Phenomenon of interest: Patient reported outcome measured with BREAST-Q, BRECON-31 and EORTC-QLQ-BRECON-23. Design: Studies examining values giving meaning to patient reported outcome measures. Evaluation: minimal important differences and normative data. Research type: All types of studies. Comments and editorials were excluded. Two of the authors (LW and EH) independently assessed if the articles met the inclusion criteria and disagreements were resolved by

discussion.

2.3. Information sources, search, and study selection

The PubMed, Medline, CINAHL, Embase, and Cochrane Library databases were searched for articles and abstracts published between January 2009 and May 2021. The first included instrument was developed 2009 and that defined the start date. The searches were performed on May 24, 2021. In addition, manuals of the instruments were scrutinized. No other grey literature searches were made. The search was limited to studies published in English, French, German, Italian, Swedish, Danish, and Norwegian. The search string was ((((((BRECON-31) OR (BREAST-Q)) OR (EORTC QLQ-BRECON-23)) OR (BRECON-23)) OR (BRECON)) OR (EORTC breast reconstruction)) AND (((((((((((((((((minimal important difference) OR (minimally important difference)) OR (MID)) OR (minimal clinically important difference)) OR (MCID)) OR (minimal detectable change)) OR (MDC)) OR (threshold value)) OR (guideline level)) OR (population norm*)) OR (normative data)) OR (normative score*)) OR (normative value*)) OR (normative standard)) OR (effect size)) OR (number needed to treat)) OR (NNT)) OR (response shift)) OR (regression to the mean)). All bibliographies of included studies were manually checked. When eligibility for inclusion could not be assessed by reading the abstract, the entire article was read and assessed.

2.4. Data collection process and data items

Information collected included country of origin, QoL/PROM instrument, sample size, response rate/loss to follow-up, demography, type of mastectomy, type of breast reconstruction, body mass index (BMI), breast size, minimal important difference, and normative data. Two authors (LW and EH) independently collected the data. Disagreements were resolved by discussion.

2.5. Assessment of risk of bias in individual studies and across studies

The included studies were critically appraised using the GRADE approach [11], rating study risk of bias (study limitations) [12], of publication bias [13], of imprecision [14], of inconsistency [15], and of indirectness [16] for the studied outcomes for individual studies as 'very low', 'low', 'moderate', 'high', or 'very high'. Overall certainty of evidence was summarised for each outcome into 'High' (ΘΘΘΘ), 'Moderate' (ΘΘΘO), 'Low' (ΘΘOO), or 'Very low' (ΘOOO) [11]. All authors independently assessed risk of bias. Disagreements were solved by discussion.

3. Results

3.1. Study selection

The literature search identified 78 articles. Of these, 34 articles were excluded after screening of abstracts. Another 41 articles were excluded when they had been read in full text (Fig. 1). Three articles were finally included in the review (Tables 1–3). The excluded articles, with reasons for exclusion, are listed in Electronic supplement 1. Two studied present normative data for BREAST-Q (Table 2) and one study MIDs for BREAST-Q (Table 3). There are no studies presenting normative data or MIDs for BRECON-31 or EORTC-QLQ-BRECON-23. All of the included studies were observational studies (Table 1).
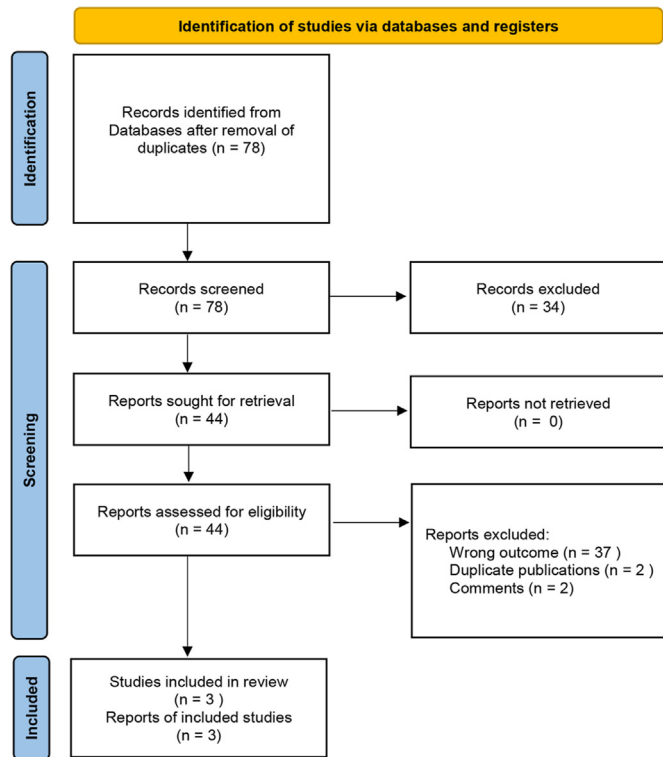
Fig. 1. PRISMA 2020 flow diagram for new systematic reviews which included searches of databases and registers only.

3.2. Risk of bias within and across studies

Regarding study limitations, the establishment of normative data does not require a control group, as a representative case series can provide high quality evidence. In the first study on normative data, the population was based on membership in the organisation Army of Women, which is an online community promoting breast cancer research. Therefore, only women who have a special interest in the promotion of breast cancer research were included in the study. Moreover, the response rate was one percent, and the demographic characteristics of the sample are not similar to that of the female American population at large. In the second study, the women were recruited among women who had an appointment in the department of gynaecology in a university hospital. The response rate is not given, and the demographic characteristics, of the sample, are different than those of the female American population at large. Hence the representativeness of the samples can be questioned. In the study on MIDs, the authors adjusted for key factors, such as type of reconstruction, radiotherapy, and weight in the analyses, but not for other factors that might affect satisfaction with breast [20], nor is there an analysis of possible baseline differences between groups. The risk of bias was evaluated as very high and moderate, respectively (Table 1).

Regarding inconsistency, the means and standard deviations of the normative scores in the two studies were similar (Table 2) and the risk of inconsistencies was evaluated as low (Table 1).

Regarding indirectness, the MIDs were calculated with distribution-based methods. There are no studies using anchor-based methods and therefore the risk of indirectness has to be

Table 1
Risk of bias across studies according to the GRADE approach [11].

First author, year, country	Design	Limitations (risk of bias)	Inconsistency	Indirectness	Imprecision	Publication bias
Klifto, 2020, USA [17]	Observational	Very high	Low	Low	Low	Low
Mundy, 2017, USA [18]	Observational	Very high	Low	Low	Low	Low
Voineskos, 2020, USA [19]	Observational	Moderate	Single study	High	Low	Low

Table 2
Normative data.

First author, year, country	Types of participants	Number of participants (response rate)	Characteristics of participants (mean (SD) or percentage)	Instrument used	Normative data Mean (SD)	Comments
Klifto, 2020, USA [17]	Participants were recruited from non-pregnant women attending a gynaecology department, with no history of breast surgery	300 (NR)	Age 43 (15) BMI: 27 (7) 42% had a bra cup size ≥ D (n = 125), 61% Caucasian (n = 182), 26% African American (n = 78), 5% Asian (n = 14), 49% were married (n = 49)	BREAST-Q	Satisfaction with breasts: 59 (21) Psychosocial wellbeing: 70 (19) Sexual well-being: 57 (19) Physical well-being (chest) 84 (13) Physical well-being (abdomen) 82 (18)	Women with increasing BMI, increasing age, Caucasian race, and bra size ≥ DD had lower scores
Mundy, 2017, USA [18]	Participants were recruited from the online community the Army of Women	1200 (1200/ 121,688, 1%)	Age: 54 (13) years BMI: 26 (6), 38% had a bra cup size ≥ D (n = 455) 92% non-Hispanic White (n = 1097), 68% were married (n = 818).	BREAST-Q	Satisfaction with breasts: 58 (18) Psychosocial wellbeing: 71 (18) Sexual well-being: 56 (18) Physical well-being (chest) 93 (11) Physical well-being (abdomen) 78 (20)	Women with BMI≥30, cup size ≥ D, age <40, and income <$40,000/year had lower scores

Table 3
Minimal important differences.

First author, year, country	Types of participants	Number of participants (response rate)	Characteristics of participants (mean (SD) or percentage)	Instrument used	Type of minimal important difference	Minimal important difference	Comments
Voineskos, 2020, USA [19]	Women who have had a mastectomy and an immediate or delayed breast reconstruction	3052		BREAST-Q	Distribution based MIDs based on 0.2 SD and 0.2 standardised response mean	Before surgery Satisfaction with breasts: 4 Psychosocial well-being: 4 Sexual well-being: 4, Physical well-being (chest): 3 1 year after surgery Satisfaction with breasts: 5 Psychosocial well-being: 4 Sexual well-being: 5 Physical well-being (chest): 3	Recommended MIDs in clinical practice Satisfaction with breasts: 4 Psychosocial well-being: 4 Sexual well-being: 4 Physical well-being (chest): 4 Recommended MIDs in clinical research: Satisfaction with breasts: 4 Psychosocial well-being: 4 Sexual well-being: 4 Physical well-being (chest): 3 Separate MIDs are given for different types of reconstruction, radiotherapy/no radiotherapy, and weight classes.

considered high (Table 1).

Regarding imprecisions, the most important factor to consider, in the present studies, is if the sample sizes were adequate. In one [17] of the included studies, the study sample was based on a generic estimation and the other two studies did not state what they based the sample size on. Nonetheless, the sample sizes of the two latter studies were 1201 and 3052 women, respectively, which has to be considered substantial for studies on breast reconstruction. Confidence intervals were not given in any of the studies. The risk of imprecisions was considered low (Table 1).

Regarding the risk of publication bias, all of the studies were conducted in North America, which could have introduced a bias. None of the studies were industry sponsored. In summary, the risk for publication bias is probably low (Table 1).

3.3. Results of individual studies and overall certainty of evidence

There are normative data for BREAST-Q based on 1500 North American women (Table 2). The overall certainty of evidence for normative data for BREAST-Q is low (GRADE ⊕⊕OO). It was down-graded due to a very high risk of bias. There are no published normative data for BRECON-31 or EORTC-QLQ-BRECON-23.

The only study on MIDs shows distribution-based MIDs, calculated, from answers from 3052 North American women, of around 4 for the subscales of BREAST-Q (Table 3). The overall certainty of evidence for MIDs for BREAST-Q is low (GRADE ⊕⊕OO). The certainty of evidence was down-graded due to a single study, a moderate risk of bias and a high risk of indirectness. There are no published MIDs for BRECON-31 or EORTC-QLQ-BRECON-23.

4. Discussion

This is a systematic review of measurements giving clinical meaning to the three PROM instruments that have adequate content validity for breast reconstruction, BREAST-Q, BRECON-31, and EORTC QLQ-BRECON-23. Three studies were found, two on normative data and one on MIDs, and all of them used BREAST-Q. Therefore, our knowledge of how PROMs evaluating breast reconstruction should be interpreted is limited.

4.1. Methodological issues

The review has some limitations, mainly concerning the evaluation of the scientific evidence of the published studies. The GRADE tool [11] was developed to assess studies on interventions and has later been developed for studies of diagnosis, prognosis, and for patient values and preferences. There are no instruments to evaluate normative data and therefore the principles of the instruments have been applied in this new field, albeit not validated.

4.2. Clinical implications and research recommendations

4.2.1. What is necessary regarding healthy controls ('normative data') for comparison?

Both studies on normative data were performed in North America. For comparisons with healthy controls to make sense, tested individuals must be representative and the instruments used have to be robustly translated and culturally validated before use [21]. This implies that specific normative data are needed for different countries and cultures, as the view of the importance of breasts and conditions for quality of life varies [22,23]. Moreover, normative data have to be accompanied by detailed characteristics regarding factors that might affect the outcome of breast reconstruction, such as socio-demography (for example age and weight) [24,25], health-state [26,27], and behaviour (for example tobacco and drug use) [20] to allow for equivalence testing [28] with specific study populations.

4.2.2. How can normative data be used? What is an acceptable result?

Other than for direct comparison of relevant groups normative data can be used to establish *threshold values* [28] for what constitutes a good and bad patient reported outcome after a breast reconstruction. If we have knowledge about how a normal population of women scores their breast satisfaction, it becomes easier to evaluate what average score is realistic to achieve with breast reconstruction and in what range the majority of patients should score. For instance, it would seem an unreasonable aim that reconstructed women are much more satisfied with their breast than matched women with natural breasts. The proportion of women scoring in different intervals, compared to the healthy

controls, could also be used to compare the outcomes of different methods and timing of reconstruction.

4.2.3. What is a clinically applicable change in scores relevant to the patient?

The result of this systematic review reveals that only MIDs based on statistical distribution are available. The establishment of valid, clinically relevant changes in scores often requires combined information from several indicators with different clinical anchors, in combination with statistical distribution [8] in different populations [29] and contexts. In breast reconstruction, it is difficult to define disease related criteria [8] as the severity of a mastectomy defect and accompanying clinical symptoms are highly subjective [30]. Moreover, the individual aim of and expectations on a breast reconstruction vary considerably [31]. Comparison to known populations [8] could be plausible as patients could be compared to patients who have a simple mastectomy and to a healthy population. Nonetheless, some people opt for a simple mastectomy and are quite happy living with the 'defect' and many healthy women are dissatisfied with their breasts [32,33], which complicates the known population concept. Another feasible option is asking the patient to rate her global change after a breast reconstruction. However, this would give a single item measure [8], not taking the complexity of patient reported outcomes after breast reconstruction into consideration. In order to establish clinically relevant differences, studies that compare changes in scores to other clinical changes ('anchors') are needed. The first challenge is to establish which anchors to use in breast reconstruction.

Furthermore, there are several other challenges in establishing clinically relevant changes in scores in breast reconstruction. The minimally clinical relevant change in score can be difficult to establish when some patients experience a limited impairment preoperatively [8], as can be the case in patients having breast reconstruction. For example, it has to be established if a greater change is required to constitute a clinically meaningful difference in patients with more severe impairment pre-operatively [8]. More extreme values are also more affected by statistical phenomena such as *regression to the mean* [8], that is a patient who is very dissatisfied preoperatively has a bigger chance of being more satisfied in a second measurement, than a patient who is moderately dissatisfied. Such factors have to be taken into consideration when clinically relevant MIDs are established for breast reconstruction.

4.2.4. When should PROMs be evaluated? When it is clinically relevant to perform the post-operative measurement?

The scores might be affected on when they are performed. For example, *response shift*, that is a change in scores over time due to changes in the patient's perception of QoL as she adapts to her condition and/or reframes her expectations [34,35]. As clinicians, we often believe that one year post-operatively is an adequate time-point to evaluate the result of a breast reconstruction, as scar maturation etc. then is complete. However, we know very little about how patients adapt to breast reconstructions over time and change their internal standards. Previous studies have revealed that, with time, QoL and psychosocial function are similar for women who have had immediate and delayed breast reconstruction and even for women who have had simple mastectomy without reconstruction [30,36], which in part could be an expression of response shift. Further studies are needed on response shift

in different timings of reconstruction and for different types of reconstruction, over time in breast reconstruction and mastectomy alone, and after breast cancer in general [37], to allow for a just comparison in clinical studies.

4.2.5. How can changes and differences in scores be used in a clinically relevant fashion? What makes one method superior to another? When are corrections worthwhile for the patient?

Many studies conclude that one technique is superior to another based on statistically significant difference between groups (e.g. Refs. [38,39]). However, a statistically significant difference does not necessary imply a clinically significant difference between groups. A discussion is needed on what magnitude of difference in scores between two methods makes one of them superior, and what magnitude of difference warrants the choice of, for example, a more expensive or resource demanding technique or a technique that implies greater risks for complications? Such an application of PROMs warrants that the data are presented in a clinically relevant way, which knowledge about scores in the healthy population and what constitutes a relevant difference in scores to patients. A more clinically relevant presentation of data could also be used in determining if minor cosmetic corrections really have an effect on patient satisfaction. Surgical corrections might be driven by a mixture of the patient's dissatisfaction and anxiety, as well as by the surgeons strive for a perfect result [40]. Nonetheless, little is known about the real impact of such corrections on long term satisfaction. The issue is of particular relevance as corrections constitute a significant part of costs and recovery in breast reconstruction [41]. Therefore, we are obliged to make sure that the procedures are truly beneficial to our patients.

Examples of more relevant ways to present results, to facilitate interpretation of the effect of different reconstructive options, could be *clinically meaningful categories* of change (improved/ deteriorated/stable/uncertain) [5,28] or the proportion of patients who have reached predefined changes in scores. Another alternative is the *number needed to treat* (*NNT*), that is the average number of patients needed to achieve one patient with improved QoL according to a predefined level, for example according to percentiles of normative data. The calculation requires that the number of patients who reach *an important change threshold* in both women who receive breast reconstruction and in those who do not is known [5,42].

5. Conclusions

There are two studies on normative data and one on MIDs of BREAST-Q and no studies of BRECON-31, and EORTC QLQ-BRECON-23. The overall certainty of evidence for normative data and MIDs for validated PROMs for breast reconstruction after mastectomy is low (GRADE ⊕⊕OO). All of the studies were performed in North America which limits the applicability in other countries and cultures. Only MIDs based on statistical characteristics are available. More studies are needed to establish relevant healthy controls and what constitutes a clinically relevant difference in scores for patient-reported outcome measures for breast reconstruction after mastectomy. Moreover, studies are needed on how results of PROM studies on breast reconstruction are best presented. Clinical implications of the findings and suggestions for further research are summarised in Fig. 2.

- Normative data from different countries and cultures with detailed information on factors, such as socio-demography, health-state, and behaviour that can affect the outcome of breast reconstruction are needed. This would allow for direct comparison with relevant controls.
- Based on normative data, threshold values and intervals for what constitutes a good and bad patient reported outcome after a breast reconstruction can be established. This would make it easier to evaluate what average score is realistic to achieve with breast reconstruction and in what range the majority of patients should score.
- Studies that compare changes in scores to other clinical changes ('anchors') are necessary to establish clinical minimal important difference in scores relevant to the patient. The first challenge is to determine relevant anchors for breast reconstruction.
- Analysis of how level of pre-operative impairment, regression to the mean, and response shift affect minimal important differences should be performed.
- Information regarding effects of timing of measurements of patient reported outcomes in breast reconstruction needs to be explored, to allow for comparison of results from different studies.
- More clinically relevant ways to present PROM results should be explored. For example, clinically meaningful categories of change (improved/deteriorated/stable/uncertain) and the number needed to treat (NNT), that is the average number of patients needed to achieve one patient with improved QoL according to a predefined level.
- A discussion is needed regarding what magnitude of difference in scores between two methods makes one of them superior, and what magnitude of difference warrants the choice of, for example, a more expensive or resource demanding technique or a technique that implies greater risks for complications.

Fig. 2. Clinical implications of the findings and suggestions for further research.

Funding

Declaration of competing interest

The authors declare that they have no competing interests.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.breast.2021.11.008.

References

[1] Potter S, Holcombe C, Ward JA, Blazeby JM, Group BS. Development of a core outcome set for research and audit studies in reconstructive breast surgery. Br J Surg 2015;102:1360–71.
[2] Potter S, Thomson HJ, Greenwood RJ, Hopwood P, Winters ZE. Health-related quality of life assessment after breast reconstruction. Br J Surg 2009;96:613–20.
[3] Sharma K, Steele K, Birks M, Jones G, Miller G. Patient-reported outcome measures in plastic surgery: an introduction and review of clinical applications. Ann Plast Surg 2019;83:247–52.
[4] Davies CF, Macefield R, Avery K, Blazeby JM, Potter S. Patient-reported outcome measures for post-mastectomy breast reconstruction: a systematic review of development and measurement properties. Ann Surg Oncol 2021;28:386–404.
[5] Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR. Clinical Significance Consensus Meeting G. Methods to explain the clinical significance of health status measures. Mayo Clin Proc 2002;77:371–83.
[6] Zhong T, Pusic AL. Future of outcomes research in plastic surgery. Clin Plast Surg 2013;40:351–7.
[7] Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Contr Clin Trials 1989;10:407–15.
[8] Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. J Clin Epidemiol 2003;56:395–407.
[9] Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005;8:19–32.
[10] Cooke A, Smith D, Booth A. Beyond PICO: the SPIDER tool for qualitative evidence synthesis. Qual Health Res 2012;22:1435–43.
[11] Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008;336:924–6.
[12] Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, et al. GRADE guidelines: 4. Rating the quality of evidence–study limitations (risk of bias). J Clin Epidemiol 2011;64:407–15.
[13] Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, et al. GRADE guidelines: 5. Rating the quality of evidence–publication bias. J Clin Epidemiol 2011;64:1277–82.
[14] Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, et al. GRADE guidelines 6. Rating the quality of evidence–imprecision. J Clin Epidemiol 2011;64:1283–93.
[15] Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 7. Rating the quality of evidence–inconsistency. J Clin Epidemiol 2011;64:1294–302.
[16] Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 8. Rating the quality of evidence–indirectness. J Clin Epidemiol 2011;64:1303–10.
[17] Klifto KM, Aravind P, Major M, Payne RM, Shen W, Rosson GD, et al. Establishing institution-specific normative data for the BREAST-Q reconstruction module: a prospective study. Aesthetic Surg J 2020;40:NP348–55.
[18] Mundy LR, Homa K, Klassen AF, Pusic AL, Kerrigan CL. Breast cancer and reconstruction: normative data for interpreting the BREAST-Q. Plast Reconstr Surg 2017;139. 1046e-55e.
[19] Voineskos SH, Klassen AF, Cano SJ, Pusic AL, Gibbons CJ. Giving meaning to differences in BREAST-Q scores: minimal important difference for breast reconstruction patients. Plast Reconstr Surg 2020;145:11e–20e.
[20] Cereijo-Garea C, Pita-Fernandez S, Acea-Nebril B, Rey-Villar R, Garcia-Novoa A, Varela-Lamas C, et al. Predictive factors of satisfaction and quality of life after immediate breast reconstruction using the BREAST-Q((c)). J Clin Nurs 2018;27:1464–74.
[21] Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR task force for translation and cultural adaptation. Value Health 2005;8:94–104.
[22] Bowen DJ, Alfano CM, McGregor BA, Kuniyuki A, Bernstein L, Meeske K, et al. Possible socioeconomic and ethnic disparities in quality of life in a cohort of breast cancer survivors. Breast Cancer Res Treat 2007;106:85–95.
[23] Fatone AM, Moadel AB, Foley FW, Fleming M, Jandorf L. Urban voices: the quality-of-life experience among women of color with breast cancer. Palliat Support Care 2007;5:115–25.
[24] Matthews H, Carroll N, Renshaw D, Turner A, Park A, Skillman J, et al. Predictors of satisfaction and quality of life following post-mastectomy breast reconstruction. Psycho Oncol 2017;26:1860–5.
[25] Srinivasa DR, Clemens MW, Qi J, Hamill JB, Kim HM, Pusic AL, et al. Obesity and breast reconstruction: complications and patient-reported outcomes in a multicenter, prospective study. Plast Reconstr Surg 2020;145. 481e-90e.
[26] Mehta SK, Olawoyin O, Chouairi F, Duy PQ, Mets EJ, Gabrick KS, et al. Worse overall health status negatively impacts satisfaction with breast reconstruction. J Plast Reconstr Aesthetic Surg 2020;73:2056–62.
[27] Mehta SK, Sheth AH, Olawoyin O, Chouairi F, Gabrick KS, Allam O, et al. Patients with psychiatric illness report worse patient-reported outcomes and receive lower rates of autologous breast reconstruction. Breast J 2020;26:1931–6.
[28] Cella D, Bullinger M, Scott C, Barofsky I. Clinical Significance Consensus Meeting G. Group vs individual approaches to understanding the clinical significance of differences or changes in quality of life. Mayo Clin Proc 2002;77:384–92.
[29] Sprangers MA, Moinpour CM, Moynihan TJ, Patrick DL, Revicki DA. Clinical Significance Consensus Meeting G. Assessing meaningful change in quality of life over time: a users' guide for clinicians. Mayo Clin Proc 2002;77:561–71.
[30] Metcalfe KA, Semple J, Quan ML, Vadaparampil ST, Holloway C, Brown M, et al. Changes in psychosocial functioning 1 year after mastectomy alone, delayed breast reconstruction, or immediate breast reconstruction. Ann Surg Oncol 2012;19:233–41.
[31] Flitcroft K, Brennan M, Spillane A. Women's expectations of breast reconstruction following mastectomy for breast cancer: a systematic review. Support Care Cancer 2017;25:2631–61.
[32] Cordeiro CN, Clarke A, White P, Sivakumar B, Ong J, Butler PE. A quantitative comparison of psychological and emotional health measures in 360 plastic surgery candidates: is there a difference between aesthetic and reconstructive patients? Ann Plast Surg 2010;65:349–53.
[33] von Soest T, Torgersen L, Kvalem IL. Mental health and psychosocial characteristics of breast augmentation patients. J Health Psychol 2020;25:1270–84.
[34] Schwartz CE, Sprangers MA. Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. Soc Sci Med 1999;48:1531–48.
[35] Sprangers MA, Schwartz CE. Integrating response shift into health-related quality of life research: a theoretical model. Soc Sci Med 1999;48:1507–15.
[36] Menez T, Michot A, Tamburino S, Weigert R, Pinsolle V. Multicenter evaluation of quality of life and patient satisfaction after breast reconstruction, a long-term retrospective study. Ann Chir Plast Esthet 2018;63:126–33.
[37] Anota A, Bascoul-Mollevi C, Conroy T, Guillemin F, Velten M, Jolly D, et al. Item response theory and factor analysis as a mean to characterize occurrence of response shift in a longitudinal quality of life study in breast cancer patients.

Health Qual Life Outcome 2014;12:32.

[38] Santosa KB, Qi J, Kim HM, Hamill JB, Wilkins EG, Pusic AL. Long-term patient-reported outcomes in postmastectomy breast reconstruction. JAMA Surg 2018;153:891—9.

[39] Fracon S, Renzi N, Manara M, Ramella V, Papa G, Arnez ZM. Patient satisfaction after breast reconstruction: implants Vs. Autologous tissues. Acta Chir Plast 2018;59:120—8.

[40] Orr JP, Sergesketter AR, Shammas RL, Thomas AB, Cason RW, Zhao R, et al. Assessing the relationship between anxiety and revision surgery following autologous breast reconstruction. Plast Reconstr Surg 2019;144:24—33.

[41] Roberts A, Baxter N, Camacho X, Lau C, Zhong T. Once is rarely enough: a population-based study of reoperations after postmastectomy breast reconstruction. Ann Surg Oncol 2015;22:3302—7.

[42] Wyrwich KW, Bullinger M, Aaronson N, Hays RD, Patrick DL, Symonds T, et al. Estimating clinically significant differences in quality of life outcomes. Qual Life Res 2005;14:285—95.