

patient-reported

**PATIENT-REPORTED
OUTCOME MEASUREMENT
GROUP, OXFORD**

**A STRUCTURED REVIEW OF
PATIENT-REPORTED OUTCOME
MEASURES FOR PATIENTS
UNDERGOING COSMETIC
SURGICAL PROCEDURES**

Report to the Department of Health 2013



O

U

T

C

O

M

E

measures

A STRUCTURED REVIEW OF PATIENT-REPORTED OUTCOME MEASURES USED IN COSMETIC SURGICAL PROCEDURES

Report to Department of Health, 2013

David Morley
Crispin Jenkinson
Ray Fitzpatrick

*Health Services Research Unit
Department of Public Health
University of Oxford*

Copies of this report can be obtained from:

David Morley
Senior Project Lead
Health Services Research Unit
Tel: +44 (0)1865 289402
e-mail: david.morley@dph.ox.ac.uk

A STRUCTURED REVIEW OF PATIENT-REPORTED OUTCOME MEASURES USED IN COSMETIC SURGICAL PROCEDURES

CONTENTS

Executive summary: 3

Chapter 1. Introduction: 4

Chapter 2. Methods: 10

Chapter 3. Cosmetic surgery specific PROMs: 12

Chapter 4. Generic PROMs utilised in cosmetic surgery: 20

Chapter 5. Summary and recommendations: 26

Appendix A. Previously published systematic reviews included in report: 29

Appendix B: Index of PROMs Included in Review: 30

Appendix C: Development, domains and scoring methods for generic PROMs: 36

Glossary of Terms: 38

References: 40

Acknowledgements: 49

EXECUTIVE SUMMARY

Aims of the report

The aim of this review is to identify patient reported outcome measures (PROMs) for use in cosmetic surgical procedures. The review assesses evidence relating to the development and validation of these instruments and make recommendations, where appropriate, regarding the most suitable instruments for use in relation to groups rather than individuals.

Methods

The primary source used to identify relevant articles was English-language PubMed records. PROMs identified were assessed on the initial development process, measurement performance and operational characteristics.

Results

The review identifies nine cosmetic surgery specific PROMs developed with patient input and that demonstrate, at least, adequate psychometric properties; the Breast Reduction Assessed Severity Scale Questionnaire (BRASSQ), Breast-Related Symptoms Questionnaire (BRSQ), BREAST-Q, Derriford Appearance Scale (DAS), Facial Lines Treatment Satisfaction Questionnaire (FTS), FACE-Q Satisfaction with Facial Appearance Scale, Patient-Reported Impact of Scars Measure (PRISM), Patient Scar Assessment Questionnaire (PSAQ) and Skindex. Three generic measures, the SF-36, EQ-5D and Healthy Utility Index, were identified as having been used for assessment of outcomes in cosmetic surgery.

Conclusions

The nine cosmetic surgery specific measures vary in their level of validation, and although some have yet to be tested for responsiveness, all demonstrate psychometric properties that justify their use. The three measures that stand out are the BREAST-Q, FACE-Q Satisfaction with Facial Appearance Scale and Skindex, all of which meet current recommendations for the development and validation of PROMs. For the three generic measures, little data was identified that reports their testing and validation in cosmetic surgery. Studies dedicated to the evaluation of these measures in this population are urgently required before they can be confidently recommended for use in the NHS.

Chapter 1: INTRODUCTION

Cosmetic Surgery in the UK

Latest figures from the British Association of Aesthetic Plastic Surgeons (BAAPS) indicate a significant rise in the number of cosmetic surgical procedures undertaken in the UK (British Association of Aesthetic Plastic Surgeons, 2012). Audit data from BAAPS for 2011 demonstrated a 5.8% increase in procedures when compared to figures from 2010, this equating to in excess of 43,000 cases.

The assessment of outcomes in cosmetic surgery is particularly pertinent, and especially from the perspective of the patient; manifestly, the patient's perception of the success, or otherwise, of their cosmetic procedure is paramount (Kosowski et al., 2009; Ching et al., 2003). Given such importance, it is essential that those PROMs utilised in the field of cosmetic surgery reflect the characteristics described in this opening chapter. The review that follows aims to identify those PROMs that are best placed for current use.

Patient Reported Outcome Measures (PROMs)

A wide range of patient-reported outcome measures (PROMs) have been developed over the last thirty years. Various terms are used to describe measures of 'health status', 'health-related quality of life', 'functional status', 'patient-reported outcome' or often just 'outcome', the common element is an attempt directly to capture the patient's experience of important aspects of health through questionnaire or interview. Considerable resources and effort have been invested to make such 'instruments' valid measures for use in relation to a wide range of decisions and policies in health. A principle problem has been that there are large numbers of such instruments from which to choose for any given health problem or context and insufficient guidance to inform choice (Garratt et al., 2002).

Such instruments generally take the form of questionnaires containing several items reflecting the broad nature of health status, disease, or injury, which are most often summed to give a total score. The term 'patient-reported outcome measure' will be used throughout this review to refer to patient-completed instruments.

There are two broad categories of PROM: generic and specific. Generic instruments are not age-, disease-, or treatment-specific and contain multiple concepts intended to be relevant to a wide range of patients and the general population. Specific instruments may be specific to a particular condition (for example, diabetes), a particular intervention or patient population. Disease-specific instruments may have greater clinical appeal due to their specificity of content, and associated increased responsiveness to specific changes in condition.

The broad content of generic instruments enables the identification of co-morbid features and unanticipated treatment side-effects that may not be captured by specific instruments, which suggests they may be useful in assessing the impact of new health-care technologies where the therapeutic effects are uncertain. However, the broad content may reduce responsiveness to small but important changes. It has therefore been recommended that a combination of generic and specific measures be used in the assessment of health outcomes.

PROMs have increasingly been applied in a range of settings including routine patient care, clinical research, audit and quality assurance, population surveys, and resource allocation. However, consensus is often lacking as to which instrument to use; this has important implications for the evaluation of clinical effectiveness. Structured reviews of measurement properties are a prerequisite for instrument selection and standardisation, and instruments with measurement properties that support their application in specific populations and across a range of evaluation settings need to be identified.

Selection criteria have been defined for assessing the quality of patient-reported health instruments (Streiner & Norman, 2008; McDowell & Newell, 2006; Fitzpatrick et al., 1998). These include measurement issues, such as reliability, validity, responsiveness, and precision, as well as practical issues, such as acceptability and feasibility. Such criteria are now regarded as essential by regulatory bodies such as the United States Food & Drug Administration (FDA). Additionally, current FDA guidance places patients at the centre of the development process of PROMs (Food & Drug Administration, Department of Health and Human Services, 2009). These criteria are now briefly summarised since they directly inform the review reported here.

Criteria for assessing PROMs

Reliability is concerned with whether measurement is accurate over time and, for multi-item instruments, whether they are internally consistent. Test-retest reliability usually involves instrument self-completion on two occasions separated by a suitable time-period and, assuming no change in the underlying health state, measures the temporal stability of the score (Fitzpatrick et al., 1998). A test-retest period of between two days and two weeks has been recommended for most conditions (Streiner & Norman, 2008). Too short a period may be associated with patient recall of answers, which may artificially inflate reliability (Nunnally & Bernstein, 1994; Streiner & Norman, 2008); too long a period may be associated with actual change in health.

Health transition questions, which invite patients to indicate whether their general or specific health has changed between instrument administrations, are often included in evaluations. This allows for the identification of stable respondents in whom intra-class correlations between scores at different administrations may be high.

The correlation coefficient is the most frequently used method for calculating estimates of test-retest reliability; the intra-class correlation coefficient (ICC) is used to identify group shift over time as a measure of reliability (Streiner & Norman, 2008). For group comparisons, levels of reliability over 0.70 are required (Streiner & Norman, 2008; Fitzpatrick et al., 1998). For the evaluation of individuals, levels above 0.90 have been recommended (Nunnally & Bernstein, 1994; Fitzpatrick et al., 1998).

Internal consistency reliability of multi-item instruments that adopt a traditional summated rating scale format is tested following a single application. The relationship between all items and their ability to measure a single underlying domain is assessed using Cronbach's alpha: alpha levels of between 0.70 and 0.90 have been recommended (Streiner & Norman, 2008; Scientific Advisory Committee of the Medical Outcomes Trust, 2002; Garratt et al., 2001). Homogeneity at the item level can be assessed using item-total correlation: levels above 0.40 have been recommended (Ware, 1997).

Validity assesses whether an instrument measures what is intended in the different settings in which it may be applied (McHorney, 1996; Fitzpatrick et al., 1998). Instrument validity is not a fixed property. The process of validity testing is on-going, informing instrument application and interpretation in different settings and with different populations (McHorney, 1996; Ware, 1997). Hence, new and refined instruments, and those applied in different settings or with different populations require evidence of validity. Both qualitative and quantitative methods can be used to assess validity.

Face and content validity require appraisal of item content, and assessment of its relationship to the instrument's proposed purpose and application (Fitzpatrick et al., 1998). Methods of item generation and instrument development may influence this assessment. Literature reviews, theoretical propositions, and interviews or focus groups with patients or health-care professionals may all inform this process. However, for patient-reported instruments to have content validity and relevance to the recipients of care, patients should be directly involved in item generation, usually via one-to-one interviews or focus groups (Fitzpatrick et al., 1998).

The quantitative assessment of validity requires comparison of the scores produced using patient-reported health instruments with those derived from other measures of health, clinical, and socio-demographic variables. Patient-reported instruments measure hypothetical constructs which are by definition non-observable, for example, HRQL and pain, and address a more general hypothesis than that supported by a specific behaviour (Nunnally & Bernstein, 1994). However, by reference to established evidence and the instrument's underlying theoretical base and item content, quantifiable relationships with a range of other instruments and clinical and socio-demographic variables can be expected (Ware, 1997; Fitzpatrick et al., 1998).

Expected correlations between variables should be presented to allow validity to be disproved (McDowell & Jenkinson, 1996). The strength of correlation between variables, be they small (less than 0.30), moderate (less than 0.50),

or large (greater than 0.70), indicates that the instrument measures the construct in a manner founded on theory or established evidence (McHorney et al., 1993). For example, two patient-reported measures of functional disability with similar content would be expected to correlate strongly. Construct validity may also be assessed using 'extreme groups', which theorises that one group will possess more or less of a construct (Streiner & Norman, 2008). For example, compared to the general older population, older people who are hospitalised following a hip fracture may be expected to report greater pain and worse HRQL.

The dimensionality or internal construct validity of a multi-item instrument can be assessed using factor analysis or principal component analysis. Principal component analysis can be used to assess the underlying structure of a multi-item instrument through the identification of components, or domains, into which items may group (McDowell, 2006). This form of analysis adds empirical weight to a hypothesised domain structure. For example, principal component analysis has supported the hypothesised eight-domain structure of the SF-36 (McHorney et al., 1993).

Responsiveness is considered a necessary measurement property of instruments intended for application in evaluative studies measuring longitudinal changes in health (Beaton et al., 2001; Liang et al., 2002). The numerous approaches to evaluating responsiveness have been reviewed by a number of authors (Liang, 1995; Wyrwich et al., 2000; Beaton et al., 2001; Liang et al., 2002; Terwee et al., 2003).

Responsiveness has been described as the ability of an instrument to measure clinically important change over time, when change is present (Fitzpatrick et al., 1998). It has also been argued that responsiveness can be viewed as longitudinal validity or as a measure of treatment effect (Terwee et al., 2003). Patient-reported health instruments have had by far the greatest application in clinical trials and most of the literature on responsiveness relates to the measurement of change in health for groups of patients (Fitzpatrick et al., 1998).

There are two broad approaches to assessing responsiveness: distribution-based and anchor-based (Wyrwich et al., 2000; Norman et al., 2001). Distribution-based approaches relate changes in instrument scores to some measure of variability, the most common method being the effect size statistic. The three widely-reported effect size statistics use the mean score change in the numerator, but have different denominators (Fitzpatrick et al., 1998). The effect size (ES) statistic uses the standard deviation of baseline scores (Liang, 1995). The standardised response mean (SRM) uses the standard deviation of the change score to incorporate the response variance in change scores. However, both the ES and SRM may be influenced by natural variance in the underlying state and by measurement error. The modified standardised response mean (MSRM), or responsiveness index, addresses the inherent natural variance that may occur in patients who otherwise report their health as unchanged, and non-specific score change by using the standard deviation of change in patients who are defined as stable (Deyo et al., 1991). In

demonstrating responsiveness to clinically important change, instruments should detect change above the non-specific change incorporated in the MSRM (Deyo et al., 1991).

It has been suggested that statistical measures of responsiveness are an insufficient basis for assessing responsiveness and that patients' views on the importance of the change should inform testing (Liang et al., 2002; Terwee et al., 2003). Anchor-based approaches assess the relationship between changes in instrument scores and an external variable (Norman et al., 2001). This includes health transition items or global judgements of change used to estimate the Minimal Important Difference (MID), the instrument change score corresponding to a small but important change (Jaeschke et al., 1989; Juniper et al., 2002). The MID can inform sample size calculations but consideration must be given to specific groups of patients and specific settings (Terwee et al., 2003). Score interpretation may be improved through the provision of evidence relating to score variation (Terwee et al., 2003) or a score range against which real change may be assessed (Streiner & Norman, 2008; Beaton et al., 2001).

External variables including transition ratings have also been compared to instrument score changes using correlation. This form of longitudinal validity (Kirshner and Guyatt, 1985; Terwee et al., 2003) assesses the extent to which changes in instrument scores concord with an accepted measure of change in patient health (Deyo et al., 1991; Fitzpatrick et al., 1998).

Precision refers to the ability of an instrument to distinguish clearly and precisely between respondents in relation to reported health or illness (Fitzpatrick et al., 1998). Ideally, items within an instrument should capture the full range of health states to be measured, supporting discrimination between respondents at clinically important levels of health (Fitzpatrick et al., 1998). Precision is influenced by several factors including response categories and item coverage of the defined concept of health purportedly measured by the instrument. Limited response categories lack precision and detail, whereas increased gradations of response increase measurement precision (Streiner & Norman, 2008; Fitzpatrick et al., 1998).

Modern psychometric methods, including Rasch analysis, are also used to assess item distribution. Where there is an uneven distribution of items across the proposed hierarchy of health, for example, item grouping in the middle range of functional ability, score change may be influenced by baseline scores and should be considered when interpreting changes in health.

Item content and response format will inevitably influence data quality and scaling, in which floor and ceiling effects are key features. Where more than 20% of responders score at the maximum level of good or bad health, score distribution generally suggests ceiling or floor effects, respectively (Streiner & Norman, 2008; Fitzpatrick et al., 1998). The greater concern is for respondents with already poor health who score at the floor of the instrument range and are consequently unable to report further deterioration in health.

Evidence suggests that floor effects are more common with instrument completion by older, sick, or disadvantaged respondents (McHorney, 1996).

Acceptability addresses the willingness or ability of patients' to complete an instrument (Fitzpatrick et al., 1998). Although difficult to evaluate directly, this is most readily assessed through instrument completion, response rates, and missing values. Where items within an instrument are consistently omitted, or difficulty is encountered in providing an answer, perhaps due to perceived irrelevance, this would suggest poor acceptability (McHorney, 1996). The font style and size used in questionnaires may also influence completion. Ideally, patients' should be interviewed for their views on instrument completion, content relevance and format during the pre-testing stage of instrument development (Fitzpatrick et al., 1998).

Reading ability is a further consideration regarding instrument acceptability (Streiner & Norman, 2008). A reading level equivalent to that of a 12 year-old has been recommended for questionnaires applicable to the general population (Streiner & Norman, 2008). However, many instruments, including the widely used Nottingham Health Profile (NHP) and the SF-36 have higher reading level requirements (McHorney, 1996; Sharples et al., 2000). It must also be remembered that reading ability may decrease with age (McHorney, 1996). Lack of familiarity with a questionnaire may further reduce response rates in older people (McHorney, 1996).

Instrument completion will also be influenced by mode of administration. Although cheaper than interview or telephone administration, postal administration often results in higher levels of missing values (McHorney, 1996; McColl et al., 2001). Evidence suggests that respondents are more willing to report less favourable health states when completing an instrument themselves than when the instrument is administered by interview (Fitzpatrick et al., 1998; Smeeth et al., 2001). Furthermore, response rates may be influenced by specific item content, for example, items relating to physical or emotional issues; the associated item relevance and appropriateness to the specific population (Bowling, 2005); and response formats, for example, visual analogue scales or Likert scaling (Fitzpatrick et al., 1998). The burden imposed by instrument length and time needed for completion is an important consideration for both respondent and clinician or researcher.

Feasibility of instrument administration refers to the time and cost of administration, scoring, and interpretation for clinicians, researchers, and other staff (Fitzpatrick et al., 1998).

Chapter 2: METHODS

Methods adopted were largely as described in previous reviews performed by the Oxford PROM group. However, due to time constraints and given the significant number of previously published reviews in this area, a strategy of updating previous reviews was adopted in part. Articles retrieved were assessed for relevance and evidence of measurement performance and operational characteristics abstracted for each PROM identified.

a) Search sources and terms

The primary source used to identify relevant articles was English-language PubMed records. Specific search terms are stated in each chapter. Supplementary searches included scanning the reference lists of review articles, checking instrument websites, where found, and drawing on other bibliographic resources.

b) Inclusion criteria

Titles and abstracts of articles were assessed for inclusion/exclusion by two independent reviewers and agreement was established. Included articles were retrieved in full. Published articles were included if they provided evidence of measurement and / or practical properties (Fitzpatrick et al., 1998) for multi-item instruments assessing aspects of health status or quality of life in patients undergoing cosmetic surgical procedures.

Specific inclusion criteria for generic and disease-specific instruments

- The instrument is patient-reported.
- There is published evidence of measurement reliability, validity or responsiveness following completion in the specified patient population.
- The instrument has been recommended for use with patients undergoing cosmetic surgical procedures.
- The instrument provides English-language versions for use among adult patients from the United Kingdom (UK), North America and Australasia.
- Evidence is available from English language publications, and instrument evaluations conducted in populations within the UK, North America and Australasia.

c) Exclusion criteria

- Clinician-assessed instruments,
- Very narrowly focused or single-item instruments
- Instruments only measuring symptoms
- Instruments without empirical evidence of measurement properties.

d) Data extraction

Data extraction followed pre-defined criteria and included both study-specific issues, such as study design and respondent characteristics, and instrument-specific issues, for example, type and description of instrument, including the domains of health status covered, length, and evidence of measurement and practical properties (McDowell 2006; Fitzpatrick et al., 1998; Garratt et al., 2002).

e) Format of the review

The summary of evidence largely follows that of previous reviews (McDowell, 2006; Fitzpatrick et al., 1998; Haywood et al., 2004). The following information is provided for each instrument:

Title

The instrument title as given by the original developer. Instrument developers, year of original publication, and any subsequent revision.

Description

The purpose and proposed application of each instrument as defined by the developers is described. Instrument development, including item derivation and number of items is summarised where available.

Measurement & practical properties

For all PROMs evidence of measurement properties (reliability, validity, responsiveness and precision) and practical properties (acceptability and feasibility) are reported in the form of tables. A tick (✓) is used to indicate that some minimal level of positive evidence was reported within the study supporting the relevant PROM.

f) PROM summaries

Although there are relatively clear cut and widely agreed criteria available to assess measurement properties of instruments, there are no clear-cut explicit criteria for how to weigh the balance of evidence or weigh the balance of evidence for instruments comparatively. The summaries reported here are based on weighing up for each of the instruments considered in detail the volume of available evidence, the quality of studies and, ultimately, the overall extent of positive and supportive evidence of measurement and practical properties. To some extent the review should be considered as based on a form of 'rapid appraisal'. It was written to inform current policy initiatives in a prompt and timely fashion. Although we are confident that we have a reasonably up-to-date and representative body of evidence to inform recommendations, in the time available it was not feasible exhaustively to search more inaccessible evidence. Nor was there time or resource to test recommendations against a consensus process of relevant user, professional and scientific judgements.

Chapter 3: Cosmetic surgery specific PROMs

The following chapter provides current information available on cosmetic surgery specific PROMs.

Search terms, results and identification of articles

Six previous reviews were adopted as the basis for this chapter. The reviews of Reavey et al, 2011; Kosowski et al, 2009; Rhee & McMullin, 2008a; 2008b; Pusic et al, 2007; Ching et al, 2003 thus formed the basis of five separate searches outlined below. The searches are summarised in Table 3.1

- **Search 1:** Based on the review of Reavey et al (2011), the term ‘patient reported outcome measures AND body contouring’ was entered into PubMed. No new relevant literature was identified since 2010.
- **Search 2:** Based on the review of Kosowski et al (2009), the term ‘patient reported outcome measures AND facial cosmetic surgery’ was entered into PubMed. One relevant article was identified since 2008.
- **Search 3:** Based on the reviews of Rhee and McMullin (2008a; 2008b), the term ‘patient reported outcome measures AND facial plastic surgery’ was entered into PubMed. Two relevant articles were identified since 2007.
- **Search 4:** Based on the review of Pusic et al (2007) the terms ‘patient reported outcome measures AND cosmetic breast surgery’ and ‘patient reported outcome measures AND reconstructive breast surgery’ were entered into PubMed. Three relevant articles identified since 2006.
- **Search 5:** Based on the review of Ching et al (2003), the term ‘patient reported outcome measures AND aesthetic surgery’ was entered into PubMed. Ten relevant articles identified since 2002.

In total the five searches identified 16 relevant articles as a means of updating previous reviews. The searches are summarised in Table 3.1

Search Number	PubMed Results	Relevant Articles
1	10	0
2	97	1
3	69	2
4	102	3
5	643	10
Total	921	16

Table 3.1: Summary of searches conducted for cosmetic surgery specific PROMs.

Of the 16 articles retained 5 articles outline the development and validation of new cosmetic surgery specific PROMs. The remaining 11 articles detail the further use and validation of such PROMs.

Measurement properties of current cosmetic surgery specific PROMs

Table 3.2 outlines the basic measurement properties of all cosmetic surgery specific PROMs detailed in both previous review articles and from the conducted searches. Those PROMs identified in previous reviews that were designed specifically and solely for use in cancer and facial paralysis patients are not included. Whilst a number of those PROMs identified in Table 3.2 report basic reliability, validity and responsiveness data, many fall short of the criteria detailed earlier. Patient input, in particular, is now regarded as essential in the development of PROMs and specifically in item generation.

Subsequently, on the basis of recommendations from previous reviews and an assessment of newly validated PROMs, nine cosmetic surgery specific PROMs were deemed appropriate for further analysis. Table 3.3 details these measures on the basis that they meet an acceptable level of the criteria outlined in Chapter 1. These PROMs might be regarded as the most 'promising' and appropriate for inclusion in future studies. A brief description of each measure, including the development process, is now provided.

Breast Reduction Assessed Severity Scale Questionnaire (BRASSQ; Sigurdson et al, 2007a;b) is a 39 item PROM, comprising five domains; Physical Implications, Poor Self-Concept, Body Pain, Negative Social Interactions and Physical Appearance. Developed through patient focus groups and expert review, the instrument demonstrates good psychometric properties through initial development and validation studies (Sigurdson et al, 2007a;b). Although described in the review of Reavey et al (2011) as 'scientifically strong', caution is required with the BRASSQ as the sample size of 101 incorporated in the validation study is small. Additionally the instrument has been criticized for its focus on physical and psychological symptoms and the absence of items relating to scarring and pain (Reavey et al, 2011).

Breast-Related Symptoms Questionnaire (BRSQ; Kerrigan et al, 2001) is a 13 item PROM developed specifically for use with patients undergoing breast reduction. The measure produces two scores. The first, the Breast Symptom Summary Score, is derived through calculation of the mean of all 13 items. The second, the Physical Symptom Count, is calculated from seven of the 13 items. Developed via patient focus groups and expert review, the instrument has acceptable psychometric properties through both the initial validation study and further development work (Kerrigan et al, 2001; Collins et al, 2002). The review of Pusic et al (2007) identified the BRSQ as the only instrument at the time that met adequate development and validation criteria. Clearly the instrument has been now superseded by those such as the BREAST-Q (Pusic et al, 2009b) and the BRASSQ (Sigurdson et al, 2007). Although criticized for a heavy emphasis on pain-related symptoms (Sigurdson et al, 2007), the BRSQ may prove attractive as a concise and simple measure to administer.

Table 3.2: PROMs previously used in the assessment of cosmetic surgical procedures

PROM Name	Validation Study	Intended Use	Number of Items	Reliability	Validity	Responsiveness
Blepharoplasty Outcomes Evaluation (BOE)	Alsarraaf et al, 2001	Blepharoplasty	6	✓	✓	✓
Breast Evaluation Questionnaire (BEQ)	Anderson et al, 2006	Breast augmentation	55	✓	✓	
Breast Reduction Assessed Severity Scale Questionnaire (BRASSQ)	Sigurdson et al, 2007a;b	Breast hypertrophy	39	✓	✓	✓
Breast-Related Symptoms Questionnaire (BRSQ)	Kerrigan et al, 2001	Breast hypertrophy	13	✓	✓	✓
Bock QoL questionnaire for patients with keloid and hypertrophic scarring.	Bock et al, 2006	Hypertrophic scars	15	✓	✓	
BREAST-Q	Pusic et al, 2009b	Breast hypertrophy	91	✓	✓	✓
Breast Chest Ratings Scale (BCRS)	Thompson & Tantleff, 1992	Breast surgery	10	✓	✓	✓
Cassileth Scar Questionnaire	Cassileth et al, 1983	Menaloma resection	12			
Dermatology Life Quality Index (DLQI)	Finlay & Khan, 1994	General dermatology, acne, rosacea, cutaneous malignancy	10	✓	✓	✓
Derriford Appearance Scale (DAS59 / DAS24)	Harris & Carr, 2001	Aesthetic surgery, facial trauma tissue repair, body image	59 / 24	✓	✓	✓
Dow Corning Questionnaire	Cash et al, 2002	Breast augmentation	18	✓		
Facelift Outcomes Evaluation (FOE)	Alsarraaf et al, 2001	Facelift	6	✓	✓	✓
FACE-Q Satisfaction with Facial Appearance Scale	Pusic et al, 2012	Facial aesthetic surgery	10	✓	✓	✓
Facial Appearance Scoring Test (FAST)	Copas & Robin, 1989	Rhinoplasty	18	✓	✓	✓
Facial Clinimetric Evaluation Scale (FaCE)	Kahn et al, 2001	Facial nerve disorders	15	✓	✓	
Facial Disability Index (FDI)	VanSwearingen & Brach, 1996	Facial nerve disorders	10	✓	✓	
Facial Lines Outcome Questionnaire (FLO)	Carruthers et al, 2002	Botox treatment of facial lines	7 / 11	✓	✓	
Facial Lines Treatment Satisfaction Questionnaire (FTS)	Cox et al, 2003	Skin rejuvenation	14	✓	✓	

Table 3.2 (continued): PROMs previously used in the assessment of cosmetic surgical procedures

PROM Name	Validation Study	Intended Use	Number of Items	Reliability	Validity	Responsiveness
Facial Injectables: Longevity, Late & Early Reactions & Satisfaction Quest (FILLERS-Q)	Sclafani et al, 2010	Facial soft tissue filler therapy	43			
Freidburg Questionnaire on Aesthetic Dermatology and Cosmetic Surgery (FQAD)	Augustin et al, 2000	General dermatology, botox treatment of facial lines	53	✓	✓	
Glasgow Benefit Inventory (GBI)	Robinson et al, 1996	Rhinoplasty botox for blepharospasm	18	✓	✓	✓
House-Brackman Facial Grading Questionnaire	Cullen et al, 2007	Facial nerve disorders	8	✓	✓	
Melasma Quality of Life Scale (MelasQOL)	Balkrishnan et al, 2003	Skin resurfacing, laser treatment	10	✓	✓	✓
Michigan Breast Reconstruction Outcomes Study – Body Image	Wilkins et al, 2000	Breast Reconstruction	9	✓		
Michigan Breast Reconstruction Outcomes Study – Satisfaction	Alderman et al, 2000	Breast Reconstruction	7	✓		
Nasal Obstruction Septoplasty Effectiveness (NOSE)	Stewart et al, 2004	Septoplasty	5	✓	✓	✓
Patient and Observer Scar Assessment Scale (POSAS)	Draaijers et al, 2004	Keloid scars, surgical scars, hypertrophic burn scars	11	✓	✓	
Patient-Rated Facial Disfigurement Analogue Scale Questionnaire	Lueg et al, 2001	Transfacial sinus surgery	2	✓	✓	
Patient-Reported Impact of Scars Measure (PRISM)	Brown et al, 2010	General scarring	37	✓	✓	
Patient Scar Assessment Questionnaire (PSAQ)	Durani et al, 2009	General scarring	39	✓	✓	
Rhinoplasty Outcomes Evaluation (ROE)	Alsarraf et al, 2001	Rhinoplasty	6	✓	✓	✓
Singh Botulinum Toxin Questionnaire	Singh et al, 2006	Botox	16	✓		
Skin Rejuvenation Outcomes Evaluation (SROE)	Alsarraf et al, 2001	Skin rejuvenation	6	✓	✓	✓
Skindex	Chren et al, 1996	General dermatology, botox, cutaneous malignancy	61 / 29 / 16	✓	✓	✓
Synkinesis Assessment Questionnaire (SAQ)	Mehta et al, 2007	Synkinesis	9	✓	✓	

Table 3.3: Cosmetic surgery specific PROMS deemed worthy of further analysis

PROM Name / Country	Development	No Items	No Dimensions	Reliability	Validity	Responsiveness	Precision	Acceptability	Feasibility
Breast Reduction Assessed Severity Scale Questionnaire (BRASSQ), Sigurdson et al, 2007. Canada	<ul style="list-style-type: none"> • Focus groups • Expert review • Validation study n = 101 	39	5	<ul style="list-style-type: none"> ✓ Internal consistency ✓ Test-retest 	<ul style="list-style-type: none"> ✓ Face ✓ Content ✓ Concurrent ✓ Construct 	✓		✓	
Breast-Related Symptoms Questionnaire (BRSQ), Kerrigan et al, 2001. US	<ul style="list-style-type: none"> • Focus groups • Expert panel • Literature review • Validation study n = 291 	13	2	<ul style="list-style-type: none"> ✓ Test-retest 	<ul style="list-style-type: none"> ✓ Face ✓ Content ✓ Construct 	✓		✓	
BREAST-Q, Pusic et al, 2009b. US / Canada	<ul style="list-style-type: none"> • Patient interview • Focus groups • Literature review • Expert panel • Cognitive interview • Validation study n = 1950 	91	6	<ul style="list-style-type: none"> ✓ Internal consistency ✓ Test-retest 	<ul style="list-style-type: none"> ✓ Face ✓ Content ✓ Concurrent ✓ Discriminant ✓ Construct ✓ Known groups 	✓	✓	✓	
Derriford Appearance Scale (DAS59 / DAS24), Harris & Carr, 2001. UK	<ul style="list-style-type: none"> • Patient interview • Literature review • Expert opinion • Validation study n = 2741 	59	5	<ul style="list-style-type: none"> ✓ Internal consistency ✓ Test-retest 	<ul style="list-style-type: none"> ✓ Face ✓ Content ✓ Concurrent ✓ Construct 	✓		✓	
FACE-Q Satisfaction with Facial Appearance Scale, Pusic et al, 2012. US / Canada	<ul style="list-style-type: none"> • Patient interview • Literature review • Expert opinion • Cognitive interview • Validation study n = 499 	10	1	<ul style="list-style-type: none"> ✓ Internal consistency 	<ul style="list-style-type: none"> ✓ Face ✓ Content ✓ Construct 	✓	✓	✓	✓
Facial LineTreat'nt Satisfaction Questionnaire (FTS), Cox et al, 2003. US	<ul style="list-style-type: none"> • Patient interview • Focus groups • Expert opinion • Validation study n = 152 	14	2	<ul style="list-style-type: none"> ✓ Internal consistency 	<ul style="list-style-type: none"> ✓ Face ✓ Content ✓ Construct ✓ Concurrent 			✓	

Table 3.3 (Continued): Cosmetic surgery specific PROMS deemed worthy of further analysis

PROM Name / Country	Development	No Items	No Dimensions	Reliability	Validity	Responsiveness	Precision	Acceptability	Feasibility
Patient-Reported Impact of Scars Measure (PRISM), Brown et al, 2010. UK	<ul style="list-style-type: none"> • Patient interview • Cognitive interview • Validation study n = 103 	37	2	<ul style="list-style-type: none"> ✓ Internal consistency ✓ Test-retest 	<ul style="list-style-type: none"> ✓ Face ✓ Content ✓ Concurrent ✓ Construct ✓ Known groups 		✓	✓	✓
Patient Scar Assessment Questionnaire (PSAQ), Durani et al, 2009. UK	<ul style="list-style-type: none"> • Patient interview • Literature review • Expert opinion • Cognitive interview • Validation study n = 667 	39	5	<ul style="list-style-type: none"> ✓ Internal consistency ✓ Test-retest 	<ul style="list-style-type: none"> ✓ Face ✓ Content ✓ Concurrent ✓ Known groups 			✓	✓
Skindex, Chren et al, 1996. US	<ul style="list-style-type: none"> • Focus groups with patients, nurses, physicians • Validation study n = 201 	61	7	<ul style="list-style-type: none"> ✓ Internal consistency ✓ Test-retest 	<ul style="list-style-type: none"> ✓ Face ✓ Content ✓ Construct ✓ Known groups 	✓		✓	✓

BREAST-Q (Pusic et al, 2009b) is a 91 item PROM for use with patients undergoing a number of breast surgery procedures including reconstruction, augmentation, hypertrophy, mastopexy, lumpectomy and mastectomy. The measure consists of six domains; Satisfaction with Breasts, Overall Outcome, Process of Care, Psychosocial Well-Being, Physical Well-Being and Sexual Well-Being. The BREAST-Q was developed via patient interviews, focus groups, literature review, expert panel and cognitive interviews. It has subsequently been subject to a rigorous validation process (Pusic et al, 2009b; Cano et al 2012; McCarthy et al, 2012), including Rasch analysis, and is therefore entirely in line with FDA requirements. For such a new measure the BREAST-Q has been widely adopted and commented upon (Macadam et al, 2012; Browne, 2012; Zhong et al, 2012; Hammond, 2012; Ward et al, 2012; Macadam et al, 2010; Chung, 2009). At 91 items the BREAST-Q takes a minimum of 10 minutes to complete and, although the authors report this as acceptable to patients, its lack of brevity may warrant further development.

Derriford Appearance Scale (DAS59 / DAS24; Harris & Carr, 2001) was developed as a 59 item PROM from which a shorter 24 item measure was later derived (Carr et al, 2005). The DAS59 was developed through a series of patient interviews, literature review and expert opinion and demonstrates excellent psychometric properties. The measure contains five dimensions; General Self-Consciousness, Social Self-Consciousness, Sexual and Bodily Self-Consciousness, Negative Self-Concept and Facial Self-Consciousness. The short-form DAS24 also demonstrates sound psychometric properties. Both the DAS59 and DAS24 can be used in patients undergoing a number of cosmetic surgical procedures, but such versatility can limit its responsiveness and some items may be less relevant to certain procedures (Reavey et al, 2011; Kosowski et al, 2009).

Facial Lines Treatment Satisfaction Questionnaire (FTS; Cox et al, 2003) is a 14 item PROM developed specifically to assess patient satisfaction with facial line treatment. The measure is comprised of two dimensions; Satisfaction with Treatment Effects and Satisfaction with Procedure. The FTS was developed through a series of patient interviews, focus groups and expert opinion, and demonstrates adequate psychometric properties. Although a well-developed measure that is quick and easy to administer, the FTS has been criticised by some for its lack of scope (Kosowski et al, 2009).

FACE-Q Satisfaction with Facial Appearance Scale (Pusic et al, 2012) is a 10 item PROM for use with patients undergoing facial aesthetic surgery. The FACE-Q project (Klassen et al, 2010) is currently on-going and aims to develop a suite of independently functioning scales designed to measure a range of important outcomes for facial aesthetic patients. The Satisfaction with Facial Appearance Scale is described as the 'core' FACE-Q scale (Pusic et al, 2012) and was developed via patient interviews, literature review, expert panel and cognitive interviews (Klassen et al, 2012). It has subsequently been subject to a rigorous validation process, including Rasch analysis, and demonstrates excellent psychometric properties. The FACE-Q Satisfaction with Facial Appearance Scale is fully compliant with FDA requirements and as a 10 item scale is likely to gain significant uptake in future research.

Patient-Reported Impact of Scars Measure (PRISM; Brown et al, 2010) is a 37 item PROM developed for use in both general practice and specialised clinics such as those for cosmetic surgery and dermatology. The measure was developed through patient and cognitive interviews and contains two dimensions; Symptoms and Quality of Life. PRISM demonstrates good psychometric properties and has been subject to Rasch Analysis. The measure has yet to be tested for responsiveness and the initial validation survey was small at 103. Further development of PRISM would enhance this promising PROM.

Patient Scar Assessment Questionnaire (PSAQ; Durani et al, 2009) is a 39 item PROM designed for use with a range of patient groups including those with cosmetic scarring. The PSAQ was subject to a rigorous development process of patient interviews, literature review, expert opinion and cognitive interview. It contains five pre-determined dimensions; Scar Appearance, Symptoms, Consciousness, Satisfaction with Scar Appearance and Satisfaction with Scar Symptoms. Whilst the measure was tested in a large sample of 667 and demonstrates promising psychometric properties, it has yet to be subject to factor analysis or Rasch analysis, and may, therefore, contain redundant items. This is acknowledged by the authors and forms part of the on-going development of the PSAQ, as does assessment of responsiveness.

Skindex (Chren et al, 1996) originated as a 61 item PROM for the assessment of QoL in patients with skin disease. Through various stages of development, both a 29 item version (Chren et al, 1997) and 16 item version (Chren et al, 2001) have subsequently been validated. Skindex has been utilised in various studies, from general dermatology through to botox treatment and cutaneous malignancy. The 61 item Skindex was developed through a process of focus groups with patients, nurses and physicians. The measure demonstrated good psychometric properties and comprised seven dimensions; Negative Affect, Self-Esteem, Anxiety, Physical Discomfort, Physical Limitations, Self-Consciousness and Intimacy. The subsequent 29 item version was validated in 508 patients and resulted in three dimensions; Emotions, Symptoms and Functioning. The measure again showed good psychometric properties. Finally, the Skindex-16 was validated in 541 patients and retained the three dimensions of the Skindex-29. Sound psychometric properties were retained. Clearly the Skindex suite of measures has been the subject of significant development and validation work. The primary objective of the authors in reducing the number of items has been to reduce respondent burden, whilst maintaining the psychometric rigour of the scales, something they appear to have achieved. The measures are widely adopted in research and the Skindex-29, in particular, has been recommended as a PROM of choice in the field (Both et al, 2007).

Details discussed in this chapter will now form part of the recommendations to be made in Chapter 5

Chapter 4: Generic PROMs Utilised in Cosmetic Surgery

The following chapter provides current information available on generic PROMs that have been incorporated in cosmetic surgery research.

Search terms, results and identification of articles

Seven generic PROMs were identified as being potentially viable for use in cosmetic surgery; the Health Utility Index (HUI; Feeny et al, 1995), Sickness Impact Profile / Functional Limitations Profile (SIP/FLP; Bergner et al, 1981), EQ-5D (EuroQol Group, 1990), Medical Outcomes Study 36-item Short Form Health Survey (SF-36; Ware & Sherbourne, 1992), Health Measurement Questionnaire (HMQ; Kind & Gudex, 1991), Dartmouth COOP Charts (Nelson et al, 1987) and Nottingham Health Profile (NHP; Hunt et al, 1981). Each PROM was searched incorporating its name and the term '.... AND aesthetic AND cosmetic surgery'. Results from the searches are summarised in Table 4.1.

PROM Name	PubMed Results	Relevant Articles
HUI	10	1
SIP/FLP	12	0
EQ-5D	14	2
SF-36	64	5
HMQ	1	0
COOP Charts	0	0
NHP	3	0

Table 4.1: Summary of searches conducted for generic PROMs utilised in cosmetic surgery

As a result of these searches the SF-36, EQ-5D and HUI are discussed in further detail. Full details of the development, domains and scoring methods of these three instruments are detailed in Appendix C.

a) SF-36:

The SF-36 was cited in 64 papers evaluating aesthetic and cosmetic surgery, and in one review, however only five provided any data evaluating the properties of the SF-36 in aesthetic or cosmetic surgery. Results are summarised in Table 4.2.

Reliability

No data available

Validity

Sigurdson et al (2007), when evaluating the validity of their new Breast Reduction Assessed Severity Scale Questionnaire (BRASSQ) compared

similar domains on their measure with similar domains on the SF-36, and found moderate though significant correlations on all relevant domains:

Physical Functioning (SF-36) and Physical Implications (BRASSQ)	0.52*
PCS (SF-36) and Physical Implications (BRASSQ)	0.45*
Bodily Pain (SF-36) and Body Pain (BRASSQ)	0.52*
Social Functioning (SF-36) and Negative Social Interactions (BRASSQ)	0.27*
Mental Health (SF-36) and Poor Self Concept (BRASSQ)	0.45*

Although the authors claim these tests were intended to provide criterion validation for the BRASSQ the results also provide construct validity for the SF-36.

Responsiveness

Klassen et al (1996a) found that the Rosenberg Self Esteem questionnaire (Rosenberg, 1965) produced greater effect sizes than any of the domains of the SF-36 in a number of surgical procedures including breast reduction, pinnaplasty, rhinoplasty and abdominoplasty. Effect sizes on the SF-36 were small or moderate, except for pain (e.s. =0.90) and physical functioning (e.s. = 0.83) in breast reduction patients. This result was supported in a further publication specifically concerned with breast reduction surgery (Klassen, et al, 1996b). Similar results were reported in a Canadian study of 57 patients undergoing reduction mammoplasty (O'Blenes, 2006).

Klassen et al (1999) report on the use of both the SF-36 and EQ-5D in patients undergoing various surgical procedures including breast reduction, rhinoplasty and abdominoplasty (n=198). For those patients indicating no change on the EQ-5D statistically significant changes were indicated for all eight dimensions on the SF-36, suggesting the SF-36 is more sensitive to change in aesthetic and cosmetic surgery than the utility index of the EQ-5D.

Precision

No data available

Acceptability

No data available

Feasibility

No data available

b) EuroQol- EQ-5D

The EQ-5D was cited in fourteen papers evaluating aesthetic and cosmetic surgery, and in one review. However, only two papers provided any data

evaluating the properties of the EuroQol in this patient group. These are summarised in Table 4.3.

Reliability

No data available

Validity

Temple-Oberle et al (2012) evaluating the construct validity of the BRECON-31, breast reconstruction satisfaction scale, against the EQ-5D in 128 women undergoing breast reconstruction. The summary scale of the BRECON-31 was found to be significantly correlated with the EQ-5D health thermometer ($r=0.50$, $p<0.01$) and utility index ($r=0.42$, $P<0.001$).

Responsiveness

Klassen et al (1999) report on the use of both the EQ-5D and SF-36 in patients undergoing various surgical procedures including breast reduction, rhinoplasty and abdominoplasty ($n=198$). For those patients indicating no change on the EQ-5D statistically significant changes were indicated for all eight dimensions on the SF-36, suggesting the EQ-5D may not be sensitive to change in cosmetic surgery.

Precision

No data available

Acceptability

No data available

Feasibility

No data available

c) Health Utility Index (HUI)

The Health Utility Index was cited in ten papers found using the search terms. Only one documented measurement properties of the HUI which is summarised in Table 4.4.

Reliability

Thoma et al (2005) report high ICCs for the HUI2 ($ICC=0.86$, no significance level indicated) and HUI3 ($ICC=0.85$, no significance level indicated). The test re-test was undertaken one week and one day prior to surgery. Thoma et al (2005) suggest the measures indicate good levels of test re-test reliability.

Validity

HUI2 and HUI3 changes scores were not correlated with changes on the SF-36 or two disease specific measures. Moderate or better correlations were reported for individual aspects of the HUI questionnaire, but as this is not how the developers recommend data from the measure be reported this is of no practical value.

Responsiveness

Change scores for HUI2 and HUI3 were not found to be significantly correlated with change scores on the SF-36 or two disease specific measures (Thoma et al, 2005). The HUI3 was found to be more responsive to change than the HUI2 (effect sizes 0.63 and 0.45 respectively). SF-36 physical and emotional summary scores were both found to be more response than HUI2. SF-36 physical component scores and change scores on two disease specific measures were substantially greater than either the HUI2 or HUI3, suggesting the HUI measures may not be as sensitive to change as competing instruments.

Minimally important differences were calculated for the HUI2 and HUI3 by Thoma et al (2005). This was the difference identified between the day before surgery and six months after surgery. For the HUI2 the MID was 0.06, which is twice the size of the MID cited by the developers (Horsman et al, 2003) and for the HUI3 it was 0.12.

Precision

No data available.

Acceptability

Of the 49 patients recruited into the study by Thoma et al (2005) 48 (98.0%) fully completed the HUI questions at baseline (one week prior to surgery), 47 (96.0%) at follow up (just prior to surgery), 42 (85.7%) at one month post op, and 43 (87.7%) and 32 (65.3%) at six months and one year post operatively. These results were similar to those gained from the SF-36 and two disease specific measures.

Feasibility

No data available.

Details discussed in this chapter will now form part of the recommendations to be made in Chapter 5

Table 4.2: Evaluation studies relating to the SF-36 in cosmetic surgery

Study/ Country	Population (N) Age Method of administration Setting	Measurement properties					
		Reliability	Validity	Responsive- ness	Precision	Acceptability	Feasibility
SF-36							
Klassen, et al, 1996a, UK	N=443 Age 32.6 (S.D. 12.3) Surgical procedures inc breast reduction, rhinoplasty, pinnaplasty, and abdominoplasty. Postal questionnaire			✓			
Klassen, et al, 1996b, UK	N=166 Age = 30.5 (S.D. 10.8, range 16-64) Patients undergoing breast reduction			✓			
Klassen, et al, 1999, UK	Surgical procedures inc breast reduction, rhinoplasty and abdominoplasty. Postal questionnaire			✓			
O'Blenes, et al, 2006	N=57 Age = 39.4 (range 21-61) Reduction mammoplasty			✓			
Sigurdson, et al 2007, Canada	N=101 Age 37.7 (range 16-66) Breast hypertrophy patients		✓ Criterion ✓ Construct				

Table 4.3: Developmental and evaluation studies relating to EQ-5D in cosmetic surgery

Study/ Country	Population (N) Age Method of administration Setting	Measurement properties					
		Reliability	Validity	Responsive- ness	Precision	Acceptability	Feasibility
EQ-5D							
Klassen, et al, 1999, UK	Surgical procedures inc breast reduction, rhinoplasty and abdominoplasty. Postal questionnaire			✓			
Temple-Oberle, et al, 2012, Canada	Breast reconstruction patients (n=128), mean age = 52.7 years. Postal questionnaire.		✓ Construct				

Table 4.4: Developmental and evaluation studies relating to the HUI in cosmetic surgery

Study/ Country	Population (N) Age Method of administration Setting	Measurement properties					
		Reliability	Validity	Responsive- ness	Precision	Acceptability	Feasibility
HUI							
Thoma et al, 2005	Breast hypertrophy patients undergoing surgical breast reduction (n=49) Age: 38 years (min=20, max=68). Questionnaires provided at clinic and either completed at the clinic, or at home and returned via the post.	✓ Test-retest	✓ Construct	✓		✓	

Chapter 5: Summary & Recommendations

This review has identified nine cosmetic surgery specific PROMs that demonstrate promising measurement properties. Three generic PROMs have also been identified that have previously been incorporated in this field, although there is little evidence that they have been adequately validated for use with cosmetic surgery patients. What is clear from the literature is that the measurement of outcomes in cosmetic surgery is still in its infancy. Early attempts at doing so were largely made with PROMs that fail to adequately meet current criteria. This, however, is changing and there is currently a move towards the development of scientifically sound cosmetic surgery specific measures.

Of the nine cosmetic surgery specific PROMs discussed in detail, all have been developed with input from patients. This is key to the development of PROMs and ensures that they reflect the concerns of the people whose outcome is being measured. The nine PROMs vary in their level of validation, and although some have yet to be tested for responsiveness, all demonstrate psychometric properties that justify their use. The three measures that stand out are the BREAST-Q, FACE-Q Satisfaction with Facial Appearance Scale and Skindex, all of which meet current recommendations for the development and validation of PROMs and are entirely in line with FDA criteria. The two scar based measures, the Patient-Reported Impact of Scars Measure and Patient Scar Assessment Questionnaire require additional development in order to establish responsiveness but both appear promising measures. Importantly, the nine instruments identified cover a range of procedures and this should allow for meaningful assessment across the wide ranging field of cosmetic surgery.

The only significant gap that there currently appears to be is a measure for the outcomes of body contouring. However, we have been made aware of the development of the BODY-Q in Canada and the USA. This questionnaire is for bariatric and body contouring surgery patients and aims to provide a PROM that can track changes in QoL from obesity through to post-body contouring surgery. The measure will be from the developers of the BREAST-Q and FACE-Q and is therefore likely to be fully FDA compliant.

With regard to generic PROMs, the review identifies three measures that have been incorporated in cosmetic surgery outcomes measurement; the SF-36, EQ-5D and Health Utility Index. There is, however remarkably little research reporting the testing and validation of these and other generic measures in cosmetic surgery. Typically, the use of measures is justified by reference to their use in other surgical groups. Consequently, it is difficult to recommend any measure with absolute certainty. Studies dedicated to the evaluation of these measures in cosmetic and aesthetic surgery are urgently needed.

It is increasingly recommended that evaluations based on PROMs should include a disease-specific and generic PROM to provide an assessment of the full range of direct

and indirect outcomes of interventions. That has been the strategy in the national PROMs programme for selected elective surgical procedures, where EQ-5D was used in combination with condition- or procedure-specific measures. However there is insufficient evidence to identify a generic measure validated for use in cosmetic surgery. Limited evidence suggests that EQ-5D may be insufficiently sensitive to patients' concerns. Whilst providing invaluable evidence for comparative purposes and for generation of health economic analyses, the EQ-5D did differ in estimates of benefits from condition-specific measures in relation to orthopaedic surgical procedures. Some further piloting of EQ-5D for cosmetic surgery may be warranted if it is considered important to be used for purposes of broad comparability with other interventions. Ideally it needs to be used in conjunction with more specific PROMs focused on cosmetic surgery. Currently no single PROM has been shown to be relevant across a spectrum of cosmetic surgical procedures.

The use of the cosmetic surgery specific PROMs discussed above may be relevant for and valuable for some specific cosmetic surgical procedures. They have been developed and validated to a standard that warrant their use for the measurement of outcomes to, at least, a satisfactory standard. However it may not be practical for an evaluative programme across cosmetic surgery to be based on diverse procedure specific PROMs. The use of generic PROMs requires further validation studies if they are to be legitimately used in the measurement of outcomes in cosmetic surgical procedures.

Appendix A: Previously published systematic reviews included in report:

Ching S, Thoma A, McCabe RE, Antony MM. (2003). Measuring outcomes in aesthetic surgery: a comprehensive review of the literature. *Plastic & Reconstructive Surgery*; 111(1): 469-480.

Kosowski TR, McCarthy C, Reavey PL, Scott AM, Wilkins EG, Cano SJ, Klassen AF, Carr N, Cordeiro PG, Pusic AL. (2009). A systematic review of patient-reported outcome measures after facial cosmetic surgery and/or nonsurgical facial rejuvenation. *Plastic & Reconstructive Surgery*; 123(6): 1819-1827.

Pusic AL, Chen CM, Cano S, Klassen A, McCarthy C, Collins ED, Cordeiro PG. (2007). Measuring quality of life in cosmetic and reconstructive breast surgery: a systematic review of patient-reported outcomes instruments. *Plastic & Reconstructive Surgery*; 120(4): 823-837.

Reavey PL, Klassen AF, Cano SJ, McCarthy C, Scott A, Rubin JP, Shermak M, Pusic AL. (2011). Measuring quality of life and patient satisfaction after body contouring: a systematic review of patient-reported outcome measures. *Aesthetic Surgery Journal*; 31(7): 807-813.

Rhee JS, McMullin BT. (2008a). Measuring outcomes in facial plastic surgery: a decade of progress. *Current Opinion in Otolaryngology & Head & Neck Surgery*; 16(4): 387-393.

Rhee JS, McMullin BT. (2008b). Outcome measures in facial plastic surgery: patient-reported and clinical efficacy measures. *Archives of Facial Plastic Surgery*; 10(3): 194-207.

Appendix B: Index of PROMs Included in Review

Blepharoplasty Outcomes Evaluation (BOE)

Alsarraf R, Larrabee WF Jr, Anderson S, Murakami CS, Johnson CM Jr. (2001). Measuring cosmetic facial plastic surgery outcomes: a pilot study. *Archives of Facial Plastic Surgery*; 3(3): 198-201.

Breast Evaluation Questionnaire (BEQ)

Anderson RC, Cunningham B, Tafesse E, Lenderking WR. (2006). Validation of the Breast Evaluation Questionnaire for use with breast surgery patients. *Plastic & Reconstructive Surgery*; 118(3): 597-602.

Breast Reduction Assessed Severity Scale Questionnaire (BRASSQ)

Sigurdson L, Kirkland SA, Mykhalovskiy E. (2007). Validation of a questionnaire for measuring morbidity in breast hypertrophy. *Plastic & Reconstructive Surgery*; 120(5): 1108-1114.

Breast-Related Symptoms Questionnaire (BRSQ)

Kerrigan CL, Collins ED, Striplin D, Kim HM, Wilkins E, Cunningham B, Lowery J. (2001). The health burden of breast hypertrophy. *Plastic & Reconstructive Surgery*; 108(6): 1591-1599.

Bock QoL questionnaire for patients with keloid and hypertrophic scarring.

Bock O, Schmid-Ott G, Malewski P, Mrowietz U. (2006). Quality of life of patients with keloid and hypertrophic scarring. *Archives of Dermatological Research*; 297(10):433-438.

BREAST-Q

Pusic AL, Klassen AF, Scott AM, Klok JA, Cordeiro PG, Cano SJ. (2009b). Development of a new patient-reported outcome measure for breast surgery: the BREAST-Q. *Plastic & Reconstructive Surgery*; 124(2): 345-353.

Breast Chest Ratings Scale (BCRS)

Thompson JK, Tantleff S. (1992). Female and male ratings of the upper torso: actual, ideal and stereotypical conceptions. *Journal of Social Behaviour & Personality*; 7, 345-354.

Cassileth Scar Questionnaire

Cassileth BR, Lusk EJ, Tenaglia AN. (1983). Patients' perceptions of the cosmetic impact of melanoma resection. *Plastic & Reconstructive Surgery*; 71(1): 73-5.

Dartmouth COOP Charts

Nelson E, Wasson J, Kirk J, Keller A, Clark D, Dietrich A, Stewart A, Zubkoff M. (1987). Assessment of function in routine clinical practice: description of the COOP Chart method and preliminary findings. *Journal of Chronic Diseases*; 40 Suppl 1: 55S-69S.

Dermatology Life Quality Index (DLQI)

Finlay AY, Khan GK. (1994). Dermatology Life Quality Index (DLQI)--a simple practical measure for routine clinical use. *Clinical & Experimental Dermatology*; 19(3): 210-216.

Derriford Appearance Scale (DAS59 / DAS24)

Harris DL, Carr AT. (2001). The Derriford Appearance Scale (DAS59): a new psychometric scale for the evaluation of patients with disfigurements and aesthetic problems of appearance. *British Journal of Plastic Surgery*; 54(3): 216-222.

Dow Corning Questionnaire

Cash TF, Duel LA, Perkins LL. (2002). Women's psychosocial outcomes of breast augmentation with silicone gel-filled implants: a 2-year prospective study. *Plastic & Reconstructive Surgery*; 109(6): 2112-2121.

EuroQol - EQ-5D

EuroQol Group. (1990). EuroQol: a new facility for the measurement of health-related quality of life. *Health Policy*; 16:199-208.

Facelift Outcomes Evaluation (FOE)

Alsarraf R, Larrabee WF Jr, Anderson S, Murakami CS, Johnson CM Jr. (2001). Measuring cosmetic facial plastic surgery outcomes: a pilot study. *Archives of Facial Plastic Surgery*; 3(3): 198-201.

FACE-Q

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 Plastic Surgery*; 26(4): 303-309.

Facial Appearance Scoring Test (FAST)

Copas JB, Robin AA. (1989). The Facial Appearance Sorting Test (FAST): an aid to the selection of patients for rhinoplasty. *British Journal of Plastic Surgery*; 42(1): 65-69.

Facial Clinimetric Evaluation Scale (FaCE)

Kahn JB, Gliklich RE, Boyev KP, Stewart MG, Metson RB, McKenna MJ. (2001). Validation of a patient-graded instrument for facial nerve paralysis: the FaCE scale. *The Laryngoscope*; 111(3): 387-398.

Facial Disability Index (FDI)

VanSwearingen JM, Brach JS. (1996). The Facial Disability Index: reliability and validity of a disability assessment instrument for disorders of the facial neuromuscular system. *Physical Therapy*; 76(12): 1288-1298.

Facial Lines Outcome Questionnaire (FLO)

Carruthers JA, Lowe NJ, Menter MA, Gibson J, Nordquist M, Mordaunt J, Walker P, Eadie N; BOTOX Glabellar Lines I Study Group. (2002). A multicenter, double-blind, randomized, placebo-controlled study of the efficacy and safety of botulinum toxin type A in the treatment of glabellar lines. *Journal of the American Academy of Dermatology*; 46(6): 840-849.

Facial Lines Treatment Satisfaction Questionnaire (FTS)

Cox SE, Finn JC, Stetler L, Mackowiak J, Kowalski JW. (2003). Development of the Facial Lines Treatment Satisfaction Questionnaire and initial results for botulinum toxin type A-treated patients. *Dermatologic Surgery*; 29(5): 444-449.

Facial Injectables: Longevity, Late and Early reactions and Satisfaction Questionnaire (FILLERS-Q)

Sclafani AP, Pizzi L, Jutkowitz E, Mueller N, Jung M. (2010). FILLERS-Q: an instrument for assessing patient experiences after treatment with facial injectable soft tissue fillers. *Facial Plastic Surgery*; 26(4): 310-319.

Freidburg Questionnaire on Aesthetic Dermatology and Cosmetic Surgery (FQAD)

Augustin M, Zschocke I, Seidenglanz K, Lange S, Schiffler A, Amon U. (2000). Validation and clinical results of the FLQA-d, a quality of life questionnaire for patients with chronic skin diseases. *Dermatology & Psychosomatics*; 1: 12-17.

Glasgow Benefit Inventory (GBI)

Robinson K, Gatehouse S, Browning GG. (1996). Measuring patient benefit from otorhinolaryngological surgery and therapy. *Annals of Otolaryngology, Rhinology & Laryngology*; 105(6): 415-422.

Health Measurement Questionnaire (HMQ)

Kind P, Gudex, C. (1991). The HMQ: measuring health status in the community. University of York, Centre for Health Economics, Health Economics Consortium.

Health Utility Index (HUI)

Feeny D, Furlong W, Boyle M, Torrance GW. (1995). Multi-Attribute Health Status Classification Systems: Health Utilities Index. *Pharmacoeconomics*; 7(6): 490-502.

House-Brackman Facial Grading Questionnaire

Cullen RD, House JW, Brackmann DE, Luxford WM, Fisher LM. (2007). Evaluation of facial function with a questionnaire: reliability and validity. *Otology & Neurotology*; 28(5): 719-722.

Melasma Quality of Life Scale (MelasQOL)

Balkrishnan R, McMichael AJ, Camacho FT, Saltzberg F, Housman TS, Grummer S, Feldman SR, Chren MM. (2003). Development and validation of a health-related quality of life instrument for women with melasma. *British Journal of Dermatology*; 149(3): 572-577.

Michigan Breast Reconstruction Outcomes Study – Body Image Questionnaire

Wilkins EG, Cederna PS, Lowery JC, Davis JA, Kim HM, Roth RS, Goldfarb S, Izenberg PH, Houin HP, Shaheen KW. (2000). Prospective analysis of psychosocial outcomes in breast reconstruction: one-year postoperative results from the Michigan Breast Reconstruction Outcome Study. *Plastic & Reconstructive Surgery*; 106(5): 1014-1025.

Michigan Breast Reconstruction Outcomes Study – Satisfaction Questionnaire

Alderman AK, Wilkins EG, Lowery JC, Kim M, Davis JA. (2000). Determinants of patient satisfaction in postmastectomy breast reconstruction. *Plastic & Reconstructive Surgery*; 106(4): 769-776.

Nasal Obstruction Septoplasty Effectiveness (NOSE)

Stewart MG, Smith TL, Weaver EM, Witsell DL, Yueh B, Hannley MT, Johnson JT. (2004). Outcomes after nasal septoplasty: results from the Nasal Obstruction Septoplasty Effectiveness (NOSE) study. *Otolaryngology – Head & Neck Surgery*; 130(3): 283-290.

Nottingham Health Profile

Hunt SM, McKenna SP, Williams J. (1981). Reliability of a population survey tool for measuring perceived health problems: a study of patients with osteoarthritis. *Journal of Epidemiology & Community Health*; 35(4): 297-300.

Patient and Observer Scar Assessment Scale (POSAS)

Draaijers LJ, Tempelman FR, Botman YA, Tuinebreijer WE, Middelkoop E, Kreis RW, van Zuijlen PP. (2004). The patient and observer scar assessment scale: a reliable and feasible tool for scar evaluation. *Plastic & Reconstructive Surgery*; 113(7): 1960-1965.

Patient-Rated Facial Disfigurement Analogue Scale Questionnaire

Lueg EA, Irish JC, Katz MR, Brown DH, Gullane PJ. (2001). A patient- and observer-rated analysis of the impact of lateral rhinotomy on facial aesthetics. *Archives of Facial Plastic Surgery*; 3(4): 241-244.

Patient-Reported Impact of Scars Measure (PRISM)

Brown BC, McKenna SP, Solomon M, Wilburn J, McGrouther DA, Bayat A. (2010). The patient-reported impact of scars measure: development and validation. *Plastic & Reconstructive Surgery*; 125(5): 1439-1449.

Patient Scar Assessment Questionnaire (PSAQ)

Durani P, McGrouther DA, Ferguson MW. (2009). The Patient Scar Assessment Questionnaire: a reliable and valid patient-reported outcomes measure for linear scars. *Plastic & Reconstructive Surgery*; 123(5): 1481-1489.

Rhinoplasty Outcomes Evaluation (ROE)

Alsarraf R, Larrabee WF Jr, Anderson S, Murakami CS, Johnson CM Jr. (2001). Measuring cosmetic facial plastic surgery outcomes: a pilot study. *Archives of Facial Plastic Surgery*; 3(3): 198-201.

Short-Form 36 (SF36)

Ware JE, Sherbourne CD. (1992). The MOS 36-Item Short-Form Health Survey (SF-36). I. Conceptual framework and item selection. *Medical Care*; 30: 473-483.

Sickness Impact Profile / Functional Limitations Profile (SIP / FLP)

Bergner M, Bobbitt RA, Carter WB, Gilson BS. (1981). The Sickness Impact Profile: development and final revision of a health status measure. *Medical Care*; 19(8): 787-805.

Singh Botulinum Toxin Questionnaire

Singh GC, Hankins MC, Dulku A, Kelly MB. (2006). Psychosocial aspects of botox in aesthetic surgery. *Aesthetic Plastic Surgery*; 30(1): 71-76.

Skin Rejuvenation Outcomes Evaluation (SROE)

Alsarraf R, Larrabee WF Jr, Anderson S, Murakami CS, Johnson CM Jr. (2001). Measuring cosmetic facial plastic surgery outcomes: a pilot study. *Archives of Facial Plastic Surgery*; 3(3): 198-201.

Skindex

Chren MM, Lasek RJ, Quinn LM, Mostow EN, Zyzanski SJ. (1996). Skindex, a quality-of-life measure for patients with skin disease: reliability, validity, and responsiveness. *Journal of Investigative Dermatology*; 107(5): 707-713.

Synkinesis Assessment Questionnaire (SAQ)

Mehta RP, WernickRobinson M, Hadlock TA. (2007). Validation of the Synkinesis Assessment Questionnaire. *The Laryngoscope*; 117(5): 923-926.

Appendix C: Development, domains and scoring methods for generic PROMs

a) SF-36: Medical Outcomes Study 36-item Short Form Health Survey (Ware and Sherbourne, 1992; Ware et al., 1994; Ware, 1997)

The Medical Outcomes Study (MOS) Short Form 36-item Health Survey (SF-36) is derived from the work of the Rand Corporation during the 1970s (Ware and Sherbourne, 1992; Ware et al., 1994; Ware, 1997). It was published in 1990 after criticism that the SF-20 was too brief and insensitive. The SF-36 is intended for application in a wide range of conditions and with the general population. Ware et al., (1994; 1997) proposed that the instrument should capture both mental and physical aspects of health. International interest in this instrument is increasing, and it is by far the most widely evaluated measure of health status (Garratt et al., 2002a).

Items were derived from several sources, including extensive literature reviews and existing instruments (Ware and Sherbourne, 1992; Ware and Gandek, 1998; Jenkinson and McGee 1998). The original Rand MOS Questionnaire (245 items) was the primary source, and several items were retained from the SF-20. The 36 items assess health across eight domains (Ware, 1997), namely bodily pain (BP: two items), general health perceptions (GH: five items), mental health (MH: five items), physical functioning (PF: ten items), role limitations due to emotional health problems (RE : three items), role limitations due to physical health problems (RP: four items), social functioning (SF: two items), and vitality (V: four items), as shown in Table 3.1. An additional health transition item, not included in the final score, assesses change in health. All items use categorical response options (range: 2-6 options). Scoring uses a weighted scoring algorithm and a computer-based programme is recommended. Eight domain scores give a health profile; scores are transformed into a scale from 0 to 100 scale, where 100 denotes the best health. Scores can be calculated when up to half of the items are omitted. Two component summary scores for physical and mental health (MPS and MCS, respectively) can also be calculated. A version of the SF-36 plus three depression questions has been developed and is variously called the Health Status Questionnaire (HSQ) or SF-36-D. The SF-36 can be self-, interview-, or telephone-administered.

b) EuroQol-EQ-5D (EuroQol Group, 1990)

The European Quality of Life instrument (EuroQol) was developed by researchers in five European countries to provide an instrument with a core set of generic health status items (The EuroQol Group, 1990; Brazier et al., 1993). Although providing a limited and standardized reflection of HRQL, it was intended that EuroQol would be supplemented by disease-specific instruments. The developers recommend the EuroQol for evaluative studies and policy research; given that health states incorporate preferences, it can also be used for economic evaluation. It can be self or interview-administered.

Existing instruments, including the Nottingham Health Profile, Quality of Well-Being Scale, Rosser Index, and Sickness Impact Profile were reviewed to inform item content

(The EuroQol Group, 1990). There are two sections to the EuroQol: the EQ-5D and the EQ thermometer. The EQ-5D assesses health across five domains: anxiety/depression (AD), mobility (M), pain/discomfort (PD), self-care (SC), and usual activities (UA), as shown in Table 3.1. Each domain has one item and a three-point categorical response scale; health 'today' is assessed. Weights based upon societal valuations of health states are used to calculate an index score of -0.59 to 1.00 , where -0.59 is a state worse than death and 1.00 is maximum well-being. A score profile can be reported. The EQ thermometer is a single 20 cm vertical visual analogue scale with a range of 0 to 100, where 0 is the worst and 100 the best imaginable health.

c) Health Utilities Index (Feeney et al, 1995)

The Health Utilities Index (HUI) was designed as a comprehensive measure of health status and health related quality of life. The Health Utilities Index (Mark 3) is a system composed of a health status classification which defines 972,000 discrete health states, and a preference, or utility, function which can be used to calculate the desirability for each health state. The HUI3 health status classification was developed by Feeney et al, (1995) to assess capacity on eight dimensions: vision, hearing, speech, ambulation, dexterity, emotion, cognition and pain/discomfort. The utility function reflects community preferences and scores each unique health state on a scale ranging from 0 (death) to 1 (perfect health). The HUI3 is a development of the Health Utilities Index containing a sub-set of items which constituted the HUI2.

Table C1: Summary of the EQ-5D, HUI and SF-36

PROM Name	Dimensions (no. items)	Response options	Score	Completion (in mins)
European Quality of Life Questionnaire (EuroQol-EQ5D) (5+1)	EQ-5D Anxiety/depression (1), Mobility (1), Pain/discomfort (1), Self-care (1), Usual activities (1) EQ-thermometer Global health (1)	EQ-5D Categorical: 3 options <i>EQ-thermometer</i> VAS Current health	EQ-5D Summation: domain profile Utility index (-0.59 to 1.00) <i>Thermometer</i> VAS (0-100)	Interview or self
Health Utility Index (Feeney et al, 1995) (8)	Vision, Hearing, Speech, Ambulation, Dexterity, Emotion, Cognition, Pain	Four domains have five response options and five have six response options	Global Utility index and single attribute utility scores for the eight separate dimensions	Self report, face to face and telephone interview
SF-36: MOS 36-item Short Form Health Survey (36)	Bodily pain (BP) (2), General health (GH) (5) Mental health (MH) (5), Physical functioning (PF) (10) Role limitation-emotional (RE) (3), Role limitation-physical (RP) (4), Social functioning (SF) (2), Vitality (V) (4)	Categorical: 2-6 options Recall: standard 4 weeks, acute 1 week	Algorithm Domain profile (0-100, 100 best health) Summary: Physical (PCS), Mental (MCS) (mean 50, sd 10)	Interview (mean values 14-15) Self (mean 12.6)

Glossary of Terms

- **Abdominoplasty:** surgical procedure to remove excess abdominal fat and skin
- **Bariatric surgery:** surgery for the treatment of weight loss
- **Blepharoplasty:** surgery of the eyelid
- **Blepharospasm:** involuntary tight contraction of the eyelid
- **Body contouring:** range of surgical procedures to remove excess skin including liposuction, brachioplasty, breast reduction abdominoplasty and thighplasty.
- **Botulinum toxin (Botox):** neurotoxin injected to reduce the appearance of wrinkles
- **Brachioplasty:** surgical procedure to remove excess fat and skin from the upper arm
- **Breast augmentation:** breast enlargement
- **Breast hypertrophy:** breast reduction
- **Cutaneous malignancy:** skin cancer
- **Dermatology:** study of the skin and its diseases
- **Hypertrophic and keloid scarring:** scars caused through the over production of collagen
- **Liposuction:** surgical procedure to remove unwanted body fat
- **Lumpectomy:** surgical procedure to remove a lump in the breast
- **Mastectomy:** surgical removal of one or both breasts
- **Mastopexy:** surgical procedure to lift the breasts
- **Melasma (also referred to as cholasma):** darkening of the skin, usually on the face
- **Otolaryngology:** study of ear nose and throat

- **Otoplasty:** surgery of the ears
- **Pinnaplasty:** correction of prominent ears
- **Rhinoplasty:** reshaping of the nose
- **Rosacea:** chronic skin condition characterised by redness, mainly on the face
- **Septoplasty:** straightening of the septum in the nose
- **Synkinesis:** involuntary movement of the face
- **Thighplasty:** surgical procedure to remove excess fat and skin from the thigh

REFERENCES

- Alderman AK, Wilkins EG, Lowery JC, Kim M, Davis JA. (2000). Determinants of patient satisfaction in postmastectomy breast reconstruction. *Plastic & Reconstructive Surgery*; 106(4): 769-776.
- Alsarraf R, Larrabee WF Jr, Anderson S, Murakami CS, Johnson CM Jr. (2001). Measuring cosmetic facial plastic surgery outcomes: a pilot study. *Archives of Facial Plastic Surgery*; 3(3): 198-201.
- Anderson RC, Cunningham B, Tafesse E, Lenderking WR. (2006). Validation of the Breast Evaluation Questionnaire for use with breast surgery patients. *Plastic & Reconstructive Surgery*; 118(3): 597-602.
- Augustin M, Zschocke I, Seidenglanz K, Lange S, Schiffler A, Amon U. (2000). Validation and clinical results of the FLQA-d, a quality of life questionnaire for patients with chronic skin diseases. *Dermatology & Psychosomatics*; 1: 12–17.
- Balkrishnan R, McMichael AJ, Camacho FT, Saltzberg F, Housman TS, Grummer S, Feldman SR, Chren MM. (2003). Development and validation of a health-related quality of life instrument for women with melasma. *British Journal of Dermatology*; 149(3): 572-577.
- Beaton DE, Bombardier C, Katz JN, Wright JG. (2001). A taxonomy for responsiveness. *Journal of Clinical Epidemiology*; 54: 1204-1217.
- Bergner M, Bobbitt RA, Carter WB, Gilson BS. (1981). The Sickness Impact Profile: development and final revision of a health status measure. *Medical Care*; 19(8): 787-805.
- Bock O, Schmid-Ott G, Malewski P, Mrowietz U. (2006). Quality of life of patients with keloid and hypertrophic scarring. *Archives of Dermatological Research*; 297(10):433-438.
- Bolton MA, Pruzinsky T, Cash TF, Persing JA. (2003). Measuring outcomes in plastic surgery: body image and quality of life in abdominoplasty patients. *Plastic & Reconstructive Surgery*; 112(2): 619-625.
- Both H, Essink-Bot ML, Busschbach J, Nijsten T. (2007). Critical review of generic and dermatology-specific health-related quality of life instruments. *Journal of Investigative Dermatology*; 127(12): 2726-2739.
- Bowling A. (2005). *Measuring Health: A Review of Quality of Life Measurement Scales* (third edition). Maidenhead: Open University Press.

Brazier JE, Jones N, Kind P. (1993). Testing the validity of the EuroQol and comparing it with the SF-36 Health Survey questionnaire. *Quality of Life Research*; 2: 169-180.

British Association of Aesthetic Plastic Surgeons. (2012). Britons tighten belts (in more ways than one). Accessed 15-11-12 <http://www.baaps.org.uk/about-us/audit/1105-britons-tighten-belts-in-more-ways-than-one>

Brown BC, McKenna SP, Solomon M, Wilburn J, McGrouther DA, Bayat A. (2010). The patient-reported impact of scars measure: development and validation. *Plastic & Reconstructive Surgery*; 125(5): 1439-1449.

Browne J. (2012). Response rates to the BREAST-Q are high in a routine practice setting. *Plastic & Reconstructive Surgery*; 130(3): 481e-482e.

Cano SJ, Klassen AF, Scott AM, Cordeiro PG, Pusic AL. (2012). The BREAST-Q: further validation in independent clinical samples. *Plastic & Reconstructive Surgery*; 129(2): 293-302.

Carr T, Moss T, Harris D. (2005). The DAS24: a short form of the Derriford Appearance Scale DAS59 to measure individual responses to living with problems of appearance. *British Journal of Health Psychology*; 10(2): 285-298.

Carruthers JA, Lowe NJ, Menter MA, Gibson J, Nordquist M, Mordaunt J, Walker P, Eadie N; BOTOX Glabellar Lines I Study Group. (2002). A multicenter, double-blind, randomized, placebo-controlled study of the efficacy and safety of botulinum toxin type A in the treatment of glabellar lines. *Journal of the American Academy of Dermatology*; 46(6): 840-849.

Cash TF, Duel LA, Perkins LL. (2002). Women's psychosocial outcomes of breast augmentation with silicone gel-filled implants: a 2-year prospective study. *Plastic & Reconstructive Surgery*; 109(6): 2112-2121.

Cassileth BR, Lusk EJ, Tenaglia AN. (1983). Patients' perceptions of the cosmetic impact of melanoma resection. *Plastic & Reconstructive Surgery*; 71(1): 73-75.

Chren MM, Lasek RJ, Sahay AP, Sands LP. (2001). Measurement properties of Skindex-16: a brief quality-of-life measure for patients with skin diseases. *Journal of Cutaneous Medicine & Surgery*; 5(2): 105-110.

Chren MM, Lasek RJ, Flocke SA, Zyzanski SJ. (1997). Improved discriminative and evaluative capability of a refined version of Skindex, a quality-of-life instrument for patients with skin diseases. *Archives of Dermatology*; 133(11): 1433-1440.

Chren MM, Lasek RJ, Quinn LM, Mostow EN, Zyzanski SJ. (1996). Skindex, a quality-of-life measure for patients with skin disease: reliability, validity, and responsiveness. *Journal of Investigative Dermatology*; 107(5): 707-713.

Chung KC. (2009). Discussion. Development of a new patient-reported outcome measure for breast surgery: the BREAST-Q. *Plastic & Reconstructive Surgery*; 124(2): 354.

Collins ED, Kerrigan CL, Kim M, Lowery JC, Striplin DT, Cunningham B, Wilkins EG. (2002). The effectiveness of surgical and nonsurgical interventions in relieving the symptoms of macromastia. *Plastic & Reconstructive Surgery*; 109(5):1556-1566.

Cox SE, Finn JC, Stetler L, Mackowiak J, Kowalski JW. (2003). Development of the Facial Lines Treatment Satisfaction Questionnaire and initial results for botulinum toxin type A-treated patients. *Dermatologic Surgery*; 29(5): 444-449.

Ching S, Thoma A, McCabe RE, Antony MM. (2003). Measuring outcomes in aesthetic surgery: a comprehensive review of the literature. *Plastic & Reconstructive Surgery*; 111(1): 469-480.

Copas JB, Robin AA. (1989). The Facial Appearance Sorting Test (FAST): an aid to the selection of patients for rhinoplasty. *British Journal of Plastic Surgery*; 42(1): 65-69.

Cullen RD, House JW, Brackmann DE, Luxford WM, Fisher LM. (2007). Evaluation of facial function with a questionnaire: reliability and validity. *Otology & Neurotology*; 28(5): 719-722.

Deyo RA, Diehr P, Patrick DL. (1991). Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. *Controlled Clinical Trials*; 12: 142S-158S.

Draaijers LJ, Tempelman FR, Botman YA, Tuinebreijer WE, Middelkoop E, Kreis RW, van Zuijlen PP. (2004). The patient and observer scar assessment scale: a reliable and feasible tool for scar evaluation. *Plastic & Reconstructive Surgery*; 113(7): 1960-1965.

Durani P, McGrouther DA, Ferguson MW. (2009). The Patient Scar Assessment Questionnaire: a reliable and valid patient-reported outcomes measure for linear scars. *Plastic & Reconstructive Surgery*; 123(5): 1481-1489.

EuroQol Group. (1990). EuroQol: a new facility for the measurement of health-related quality of life. *Health Policy*; 16:199-208.

Feeny D, Furlong W, Boyle M, Torrance GW. (1995). Multi-Attribute Health Status Classification Systems: Health Utilities Index. *PharmacoEconomics*; 7(6): 490-502.

Finlay AY, Khan GK. (1994). Dermatology Life Quality Index (DLQI)--a simple practical measure for routine clinical use. *Clinical & Experimental Dermatology*; 19(3): 210-216.

Fitzpatrick R, Davey C, Buxton MJ, Jones DR. (1998). Evaluating patient-based outcome measures for use in clinical trials. *Health Technology Assessment*; 2(14).
Food & Drug Administration, Department of Health and Human Services. (2009). *Guidance for Industry. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labelling Claims*. Silver Spring, MD: Food and Drug Administration.

Garratt AM, Schmidt L, Mackintosh A, Fitzpatrick R. (2002). Quality of life measurement: bibliographic study of patient assessed health outcome measures. *British Medical Journal*; 324(7351): 1417-1421.

Garratt AM, Hutchinson A, Russell I. (2001) The UK version of the Seattle Angina Questionnaire (SAQ-UK): reliability, validity and responsiveness. *Journal of Clinical Epidemiology*; 54: 907-915.

Harris DL, Carr AT. (2001). The Derriford Appearance Scale (DAS59): a new psychometric scale for the evaluation of patients with disfigurements and aesthetic problems of appearance. *British Journal of Plastic Surgery*; 54(3): 216-222.

Hammond DC. (2012). Discussion. The BREAST-Q: further validation in independent clinical samples. *Plastic & Reconstructive Surgery*; 129(2): 303-304.

Haywood KL, Garratt AM, Schmidt L, Mackintosh AE, Fitzpatrick R. (2004). *Health Status and Quality of Life in Older People: a Structured Review of Patient-assessed Health Instruments* Report from the Patient-assessed Health Instruments Group to the Department of Health.

Horsman J, Furlong B, Feeny D, Torrance G. (2003). The Health Utilities Index (HUI): Concepts, measurement properties and applications. *Health & Quality of Life Outcomes*; 1: 54.

Hunt SM, McKenna SP, Williams J. (1981). Reliability of a population survey tool for measuring perceived health problems: a study of patients with osteoarthritis. *Journal of Epidemiology & Community Health*; 35(4): 297-300.

Jaeschke R, Singer J, Guyatt GH. (1989). Measurement of health status. Ascertaining the minimal clinically important difference. *Controlled Clinical Trials*; 10: 407-415.

Jenkinson C, McGee H. (1998). *Health Status Measurement, a Brief but Critical Introduction*. Radcliffe Medical Press Ltd, Oxford.

Juniper EF, Price DB, Stampone PA, Creemers JP, Mol SJ, Fireman P. (2002). Clinically important improvements in asthma-specific quality of life, but no difference in

conventional clinical indexes in patients changed from conventional beclomethasone dipropionate to approximately half the dose of extrafine beclomethasone dipropionate. *Chest*; 121(6): 1824-1832.

Kahn JB, Gliklich RE, Boyev KP, Stewart MG, Metson RB, McKenna MJ. (2001). Validation of a patient-graded instrument for facial nerve paralysis: the FaCE scale. *The Laryngoscope*; 111(3): 387-398.

Kerrigan CL, Collins ED, Striplin D, Kim HM, Wilkins E, Cunningham B, Lowery J. (2001). The health burden of breast hypertrophy. *Plastic & Reconstructive Surgery*; 108(6): 1591-1599.

Kind P, Gudex, C. (1991). The HMQ: measuring health status in the community. University of York, Centre for Health Economics, Health Economics Consortium.

Kirshner B, Guyatt G. (1985). A methodological framework for assessing health indices. *Journal of Chronic Diseases*; 38: 27-36.

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 Plastic Surgery*; 26(4): 303-309.

Klassen AF, Fitzpatrick R, Jenkinson C, Goodacre T. (1999). Contrasting evidence of the effectiveness of cosmetic surgery from two health related quality of life measures. *Journal of Epidemiology & Community Health*; 53: 440-441.

Klassen AF, Jenkinson C, Fitzpatrick R, Goodacre T. (1996a). Patient's health related quality of life before and after aesthetic surgery. *British Journal of Plastic Surgery*; 49: 433-438.

Klassen AF, Fitzpatrick R, Jenkinson C, Goodacre T. (1996b). Should breast reduction surgery be rationed? A comparison of the health status of patients before and after treatment: postal questionnaire survey. *BMJ*; 313: 454-457.

Kosowski TR, McCarthy C, Reavey PL, Scott AM, Wilkins EG, Cano SJ, Klassen AF, Carr N, Cordeiro PG, Pusic AL. (2009). A systematic review of patient-reported outcome measures after facial cosmetic surgery and/or nonsurgical facial rejuvenation. *Plastic & Reconstructive Surgery*; 123(6): 1819-1827.

Liang MH, Lew RA, Stucki G, Fortin PR, Daltroy L. (2002). Measuring clinically important changes with patient-oriented questionnaires. *Medical Care*. 40(4): Supp II45-II51.

Liang MH. (1995). Evaluating measurement responsiveness. *The Journal of Rheumatology*; 22(6): 1191-1192.

Lueg EA, Irish JC, Katz MR, Brown DH, Gullane PJ. (2001). A patient- and observer-rated analysis of the impact of lateral rhinotomy on facial aesthetics. *Archives of Facial Plastic Surgery*; 3(4): 241-244.

Macadam SA, Ho AL, Lennox PA, Pusic AL. (2012). Patient-Reported Satisfaction and Health Related Quality of Life Following Breast Reconstruction: A Comparison of Shaped Cohesive Gel and Round Cohesive Gel Implant Recipients. *Plastic & Reconstructive Surgery*; Epub ahead of print.

Macadam SA, Ho AL, Cook EF Jr, Lennox PA, Pusic AL. (2010). Patient satisfaction and health-related quality of life following breast reconstruction: patient-reported outcomes among saline and silicone implant recipients. *Plastic & Reconstructive Surgery*; 125(3):761-771.

McCarthy CM, Cano SJ, Klassen AF, Scott A, Van Laeken N, Lennox PA, Cordeiro PG, Pusic AL. (2012). The magnitude of effect of cosmetic breast augmentation on patient satisfaction and health-related quality of life. *Plastic & Reconstructive Surgery*; 130(1): 218-223.

McColl E, Jacoby A, Thomas L, Soutter J, Bamford C, Steen N, Thomas R, Harvey E, Garratt A, Bond J. (2001). Design and use of questionnaires: a review of best practice applicable to surveys of health service staff and patients. *Health Technology Assessment*; 5(31):1-256

McDowell I. (2006). *Measuring Health: a Guide to Rating Scales & Questionnaires* (Third Edition). New York: Oxford University Press.

McDowell I, Jenkinson C. (1996). Development standards for health measures. *Journal of Health Service Research Policy*; 1(4): 238-246.

McHorney CA. (1996). Measuring and monitoring general health status in elderly persons: practical and methodological issues in using the SF-36 Health Survey. *Gerontologist*; 36: 571-583.

McHorney CA, Ware JE, Raczek AE. (1993). The MOS-36-item Short-Form Health Survey (SF-36): II. Psychometric and clinic tests of validity in measuring physical and mental health constructs. *Medical Care*; 31(3): 247-263.

Mehta RP, WernickRobinson M, Hadlock TA. (2007). Validation of the Synkinesis Assessment Questionnaire. *The Laryngoscope*; 117(5): 923-926.

Nelson E, Wasson J, Kirk J, Keller A, Clark D, Dietrich A, Stewart A, Zubkoff M. (1987). Assessment of function in routine clinical practice: description of the COOP Chart method and preliminary findings. *Journal of Chronic Diseases*; 40 Suppl 1: 55S-69S.

Norman GR, Sridhar FG, Guyatt GH, Walter SD. (2001). Relation of distribution and anchor-based approaches in interpretation of changes in health-related quality of life. *Medical Care*; 39(10): 1039-1047.

Nunnally JC, Bernstein IH. (1994). *Psychometric Theory* (Third Edition). McGraw-Hill Series in Psychology, McGraw-Hill, Inc.

O'Blenes CA, Delbridge CL, Miller BJ, Pantelis A, Morris SF. (2006). Prospective study of outcomes after reduction mammoplasty: long-term follow-up. *Plastic & Reconstructive Surgery*; 117(2): 351-358.

Pusic AL, Klassen AF, Scott AM, Cano SJ. (2012). Development and Psychometric Evaluation of the FACE-Q Satisfaction with Appearance Scale: A New PRO Instrument for Facial Aesthetics Patients. *Clinics in Plastic Surgery*; In Press.

Pusic AL, Reavey PL, Klassen AF, Scott A, McCarthy C, Cano SJ. (2009a). Measuring patient outcomes in breast augmentation: introducing the BREAST-Q Augmentation module. *Clinics in Plastic Surgery*; 36(1): 23-32.

Pusic AL, Klassen AF, Scott AM, Klok JA, Cordeiro PG, Cano SJ. (2009b). Development of a new patient-reported outcome measure for breast surgery: the BREAST-Q. *Plastic & Reconstructive Surgery*; 124(2): 345-353.

Pusic AL, Chen CM, Cano S, Klassen A, McCarthy C, Collins ED, Cordeiro PG. (2007). Measuring quality of life in cosmetic and reconstructive breast surgery: a systematic review of patient-reported outcomes instruments. *Plastic & Reconstructive Surgery*; 120(4): 823-837.

Reavey PL, Klassen AF, Cano SJ, McCarthy C, Scott A, Rubin JP, Shermak M, Pusic AL. (2011). Measuring quality of life and patient satisfaction after body contouring: a systematic review of patient-reported outcome measures. *Aesthetic Surgery Journal*; 31(7): 807-813.

Rhee JS, McMullin BT. (2008a). Measuring outcomes in facial plastic surgery: a decade of progress. *Current Opinion in Otolaryngology & Head & Neck Surgery*; 16(4): 387-393.

Rhee JS, McMullin BT. (2008b). Outcome measures in facial plastic surgery: patient-reported and clinical efficacy measures. *Archives of Facial Plastic Surgery*; 10(3): 194-207.

Robinson K, Gatehouse S, Browning GG. (1996). Measuring patient benefit from otorhinolaryngological surgery and therapy. *Annals of Otolaryngology, Rhinology & Laryngology*; 105(6): 415-422.

Rosenberg M. (1965) *Society and the Adolescent Self Image*. Princeton: Princeton University Press.

Scientific Advisory Committee of the Medical Outcomes Trust (2002). Assessing health status and quality of life instruments: Attributes and review criteria. *Quality of Life Research*; 11, 193-205.

Sclafani AP, Pizzi L, Jutkowitz E, Mueller N, Jung M. (2010). FILLERS-Q: an instrument for assessing patient experiences after treatment with facial injectable soft tissue fillers. *Facial Plastic Surgery*; 26(4): 310-319.

Sharples LD, Todd CJ, Caine N, Tait S. (2000). Measurement properties of the Nottingham Health Profile and Short Form 36 health status measures in a population sample of elderly people living at home: results from ELPHS. *British Journal of Health Psychology*; 5: 217-233.

Sigurdson L, Kirkland SA, Mykhalovskiy E. (2007b). Validation of a questionnaire for measuring morbidity in breast hypertrophy. *Plastic & Reconstructive Surgery*; 120(5): 1108-1114.

Sigurdson L, Mykhalovskiy E, Kirkland SA, Pallen A. (2007a). Symptoms and related severity experienced by women with breast hypertrophy. *Plastic & Reconstructive Surgery*; 119(2): 481-486.

Singh GC, Hankins MC, Dulku A, Kelly MB. (2006). Psychosocial aspects of botox in aesthetic surgery. *Aesthetic Plastic Surgery*; 30(1): 71-76.

Smeeth L, Fletcher AE, Stirling S, Nunes M, Breeze E, Ng E, Bulpitt CJ, Jones D. (2001). Randomised comparison of three methods of administering a screening questionnaire to elderly people: findings from the MRC trial of the assessment and management of older people in the community. *British Medical Journal*; 323: 1403-1407.

Stewart MG, Smith TL, Weaver EM, Witsell DL, Yueh B, Hannley MT, Johnson JT. (2004). Outcomes after nasal septoplasty: results from the Nasal Obstruction Septoplasty Effectiveness (NOSE) study. *Otolaryngology – Head & Neck Surgery*; 130(3): 283-290.

Streiner DL, Norman GR. (2008). *Health Measurement Scales. A practical guide to their development and use (Fourth Edition)*. Oxford: Oxford University Press.

Temple-Oberle C, Cook EF, Bettger-Hahn M, Mychailyshyn N, Naeem H, MacDermid J. (2012) Development of a Breast reconstruction Satisfaction Questionnaire (BRECON-31): Principal components analysis and clinimetric properties. *Journal of Surgical Oncology*; 106: 799-806.

Terwee CB, Dekker FW, Wiersinga WM, Prummel MF, Bossuyt PMM. (2003). On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. *Quality of Life Research*; 12: 349-362.

Thoma A, Sprague S, Veltri K, Duku E, Furlong W. (2005). Methodology and measurement properties of health-related quality of life instruments: A prospective study of patients undergoing breast reduction surgery. *Health & Quality of Life Outcomes*; 3: 44

Thompson JK, Tantleff S. (1992). Female and male ratings of the upper torso: actual, ideal and stereotypical conceptions. *Journal of Social Behaviour & Personality*; 7, 345-354.

VanSwearingen JM, Brach JS. (1996). The Facial Disability Index: reliability and validity of a disability assessment instrument for disorders of the facial neuromuscular system. *Physical Therapy*; 76(12): 1288-1298.

Ward JA, Potter S, Blazeby JM; BRAVO Study Steering Committee. (2012). The BREAST-Q: further validation in independent clinical samples. *Plastic & Reconstructive Surgery*; 130(4): 617e-618e.

Ware JE, Sherbourne CD. (1992). The MOS 36-Item Short-Form Health Survey (SF-36). I. Conceptual framework and item selection. *Medical Care*; 30: 473-483.

Ware J, Kosinski M, Keller SD. (1994). SF-36 Physical and Mental Health Summary Scales: A User's Manual. Boston, MA: The Health Institute, New England Medical Centre.

Ware JE. (1997). SF-36 Health Survey. Manual and Interpretation Guide (Second Edition). The Health Institute, New England Medical Centre. Boston, MA. Nimrod Press.

Ware JE, Gandek B. (1998). Methods for testing data quality, scaling assumptions and reliability: the IQOLA project approach. *Journal of Clinical Epidemiology*; 51(11): 945-952.

Wilkins EG, Cederna PS, Lowery JC, Davis JA, Kim HM, Roth RS, Goldfarb S, Izenberg PH, Houin HP, Shaheen KW. (2000). Prospective analysis of psychosocial outcomes in breast reconstruction: one-year postoperative results from the Michigan Breast Reconstruction Outcome Study. *Plastic & Reconstructive Surgery*; 106(5): 1014-1025.

Wyrwich KW, Wolinsky FD. (2000). Identifying meaningful intra-individual change standards for health-related quality of life measures. *Journal of Evaluation in Clinical Practice*; 6(1): 39-49.

Zhong T, McCarthy C, Min S, Zhang J, Beber B, Pusic AL, Hofer SO. (2012). Patient satisfaction and health-related quality of life after autologous tissue breast reconstruction: a prospective analysis of early postoperative outcomes. *Cancer*, 118(6): 1701-1709.

Acknowledgements: We would like to acknowledge the assistance of Dr Anne Klassen, McMaster University, Canada.